UNIVERSITY OF CALIFORNIA
RIVERSIDE

Investigations into the Cytotoxic and Mutagenic Effects of DNA Adducts and DNA Structures on DNA Replication

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Nisana Marie Andersen

December 2013

Dissertation Committee:
Dr. Yinsheng Wang, Chairperson
Dr. Ryan R. Julian
Dr. Dallas L. Rabenstein

The Dissertation of Nisana Marie Andersen is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

This material is based on work supported by the National Institutes of Health under Grant R01 DK082779 and by the Tobacco-Related Disease Research Program Grant 19DT-0009. Any opinions, findings, and conclusions or recommendations expressed in this material are my own, and do not necessarily reflect the views of the National Institutes of Health or the Tobacco-Related Disease Research Program.

COPYRIGHT ACKNOWLEDGEMENTS

The text and figures in Chapter 1 and Appendix A, in part or in full, are a reprint of the material as it appears in ***Chemical Research in Toxicology 2012***, 25, p. 2523-2531 and the supporting information therein. The co-author (Dr. Yinsheng Wang) listed in this publication directed and supervised the research which forms the basis for this chapter. The co-authors (Dr. Jianshuang Wang and Pengcheng Wang) listed in this publication synthesized the methylated phosphoramidites used to construct the damaged containing templates. The co-author (Dr. Yong Jiang) listed in this publication purified the yeast polymerase η used in the replication experiments.

The text and figures in Chapter 3 and Appendix C, in part or in full, are a reprint of the material as it appears in ***The Journal of Biological Chemistry***, 286, p. 17503-17511 and the supporting information therein. The co-author (Dr. Yinsheng Wang) listed in this publication directed and supervised the research which forms the basis for this chapter. The co-author (Dr. Bifeng Yuan) listed in this publication performed the *in vivo* mammalian replication experiments. The co-author (Dr. Yong Jiang) listed in this publication purified the yeast polymerase η used in the replication experiments.

DEDICATION

This work is dedicated to my mom, dad and husband. Thank you for all of your love and support over the years.

ABSTRACT OF THE DISSERTATION

Investigations into the Cytotoxic and Mutagenic Effects of DNA Adducts and Endogenous DNA Structures on DNA Replication

by

Nisana Marie Andersen

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, December 2013
Dr. Yinsheng Wang, Chairperson

Byproducts of normal cellular processes and environmental toxicants are capable of interacting with cellular DNA, producing a variety of chemical modifications. In addition, when duplex DNA becomes single-stranded, certain DNA sequences are capable of adopting non-B structures potentially causing genomic instability. To avoid activation of cellular checkpoints due to stalled replication forks, cells are equipped with translesion synthesis (TLS) polymerases capable of bypassing DNA lesions arising from these endogenous and exogenous sources. TLS polymerases are able to carry out lesion bypass because they lack the proofreading activity of replicative polymerases and possess larger active sites. In some cases lesion bypass carried out by these polymerases is accurate and efficient, while in other cases, TLS introduces mutations into the genome. Investigations into the cytotoxic and mutagenic effects DNA lesions and non-B structures have on DNA replication are needed to understand their role in cancer, aging and disease.

In these studies, comprehensive experiments were conducted to examine the effect regioisomeric alkylated thymidine lesions have on DNA replication; specifically the role TLS polymerases play in processing these lesions. Produced as a result from exposure to tobacco smoke, alkylated thymidines were shown to be poorly repaired in cells and thus, likely to contribute the mutations detected in lung cancer patients. Using novel LC-MS/MS methods, alkylated thymidines were shown to be both blocking and highly mutagenic to most DNA polymerases *in vitro*, primarily introducing T→C and T→A mutations. On the other hand, endogenously produced N^2-alkyl-2'-deoxyguanosine lesions were shown to be accurately and efficiently bypassed through the combination of two TLS polymerases; pol κ or pol ι inserts the correct dCMP opposite the lesion, and pol ζ extends past the lesion. In addition, cells deficient in pol κ and pol ι showed elevated levels of G→A and G→T mutations, which were attributed to TLS carried out by pol η.

Taking advantage of the competitive replication and adduct bypass (CRAB) assay, we also investigated the effect non-B, G-quadruplex (G4) structures have on DNA replication. The experiments revealed that two G4 sequences located in the promoter regions of the oncogene *c-Kit* and the proto-oncogene *c-Myb* were capable of blocking DNA replication. Additionally, *E. coli* TLS pols II and IV, along with *E. coli* DNA helicase RecD, were shown to be involved in resolving the *c-Kit*-G4 sequence.

Table of Contents

Chapter 3: Replication Across Regioisomeric Ethylated Thymidine Lesions by Purified DNA Polymerases

List of Schemes

List of Figures

List of Tables

Chapter 1: Introduction

Overview of DNA damage

Cellular DNA is susceptible to damage from a variety of exogenous and endogenous agents.[1] Examples of exogenous sources of DNA damage include, UV irradiation,[2, 3] γ radiation,[4] natural toxins,[5] and organic and inorganic environmental toxicants.[6] Reactive oxygen species (ROS) generated endogenously as byproducts of normal cellular metabolism and spontaneous hydrolysis of DNA bases can also lead to DNA damage.[7, 8] These exogenous and endogenous damaging agents can lead to a variety of lesions including alkylation and oxidation of nucleobases, intra- and inter-strand crosslinks, deamination, depurination and depyrimidination, as well as DNA strand breaks.[7, 8] With an estimated one million DNA bases damaged per cell per day, it is important to characterize how those damaged bases affect DNA replication.

Formation of Alkylated DNA from Tobacco Smoke

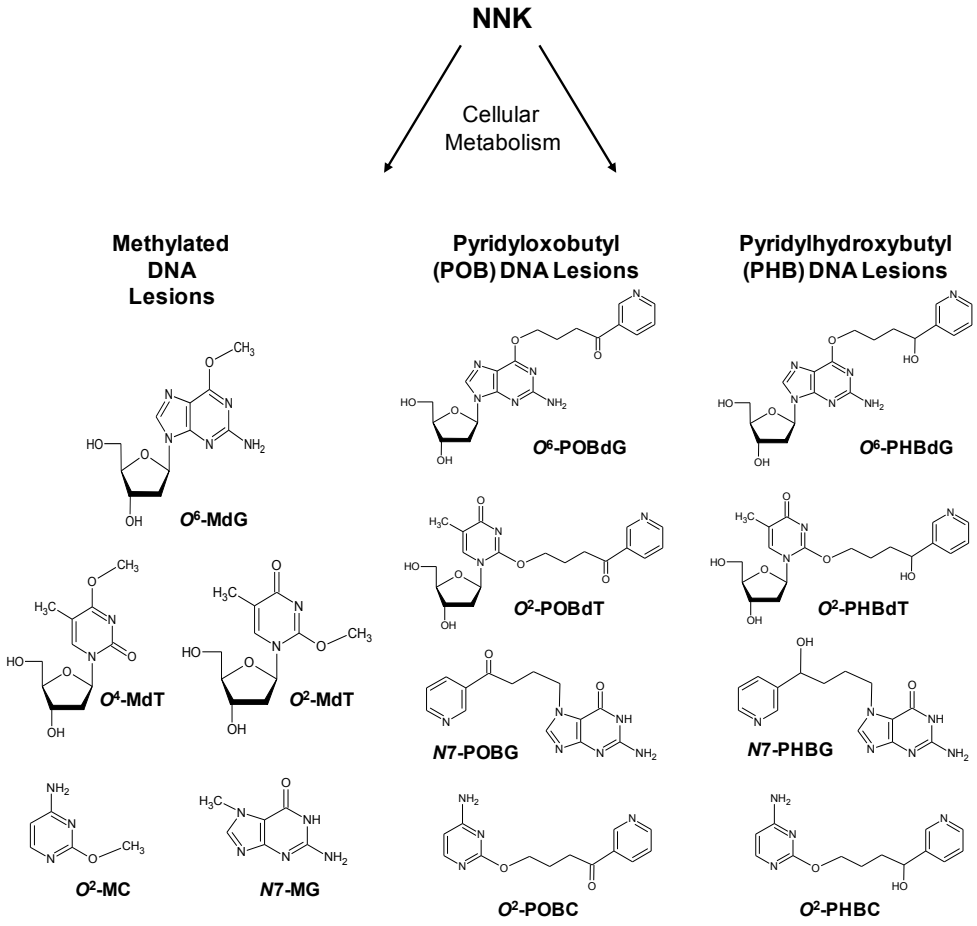
One important source of DNA damage is the consumption of tobacco products. It is estimated that over 50% of males and 10% of females smoke cigarettes worldwide.[9] In addition, one in five deaths in the United States can be attributed to smoking; smoking-induced human diseases range from lung cancer, coronary heart disease, chronic obstructive pulmonary disease and other airway obstructions.[10] Lung cancer, and heart and pulmonary disease progressively develop from damage by chemicals to DNA in cells. Tobacco smoke contains over 4000 chemicals, with approximately 40 identified as

carcinogenic to humans and animals.[11-13] These carcinogenic chemicals include acetaldehyde, which is capable of directly interacting with DNA bases, while others such as tobacco-specific nitrosamines require metabolic activation before they are able to exert their toxicity.

Metabolic activation of tobacco-specific nitrosamines can lead to a variety of alkylated DNA bases. 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone) (NNK) and *N'*-nitrosonornicotine (NNN) are major tobacco-specific nitrosamines, and they are considered carcinogenic to animals and humans.[11] In particular, NNK is capable of producing both small and large alkylated DNA adducts. Small, methylated DNA adducts include O^6-methyl-2'-deoxyguanosine (O^6-MdG), O^4-methylthymidine (O^4-MdT) and *N*7-methyl-2'-guanine (*N7*-MG); however, other methylation products including O^2-methylthymidine (O^2-MdT) and O^2-methylcytosine (O^2-MC) are also formed.[14-16] Larger, pyridyloxobutyl (POB) and pyridylhydroxybutyl (PHB) derivatives on the O^2 position of cytosine and thymine bases as well as the *N*7 and O^6 positions of guanine are also produced.[17-20] On the other hand, metabolism of NNN leads to the formation of alkyl products at the N^2 position of dG.[21] Scheme 1.1 shows the DNA adducts produced from the metabolism of NNK and NNN.

In addition to the formation of methylated, pyridyloxobutylated and pyridylhydroxybutylated DNA lesions, tobacco smoke contains a direct-acting ethylating agent that has yet to be identified, capable of transferring an ethyl group to DNA bases.[22] *In vitro* exposure to cigarette smoke resulted in a dose-dependent increase in the level of *N*7-ethylguanine (*N*7-EtG). In the same vein, the levels of *N*3-ethyladenine (*N*3-EtA) [6, 23]

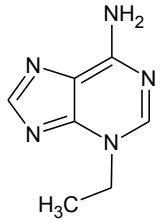
Scheme 1.1. Metabolic activation of NNK and the formation of DNA adducts



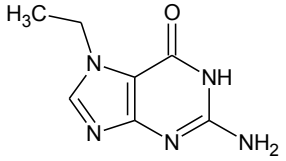
and *N*7-EtG in urine,[24] O^4-ethylthymidine (O^4-EtdT) in lung tissue,[25] *N*7-EtG in human leukocyte DNA,[26] and O^2-, *N*3- and O^4-ethylthymidine (O^2-EtdT, *N*3-EtdT, and O^4-EtdT) in human leukocyte DNA[27] were significantly higher in smokers than non-smokers. The structures of the ethylated DNA lesions in human leukocytes discussed above are sumarized in Figure 1.1.

Understanding the mechanisms by which tobacco smoke induces cancer is of utmost importance. From the aforementioned studies it is clear that DNA adduct formation can be linked to tobacco smoke exposure. If the resulting DNA adduct is not rapidly repaired, depending upon the nature of the adduct, it can lead to mutations, or other adverse cellular events. Of particular interest are alkylated thymidines, whose biological properties have not been well studied. In NNK-treated A/J mice and rats, the highly mutagentic O^6-POB-dG adduct was present at very low levels, whereas the O^2-POB-dT adduct was amongst the most persistent.[28-30] In addition, Chinese hamster ovary (CHO) cells treated with 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone (NNKOAc) primarily produced point mutations at A:T base pairs, suggesting a mutagenic potential of alkylated thymidine lesions.[31] Together, these studies underscore a potential mutagenic role for alkylated thymidines.

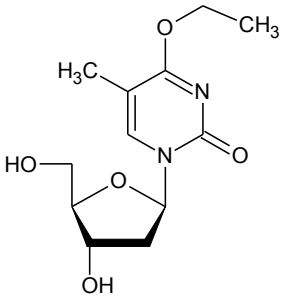
Figure 1.1. Common ethylated DNA adducts detected at elevated levels in smokers compared to nonsmokers.



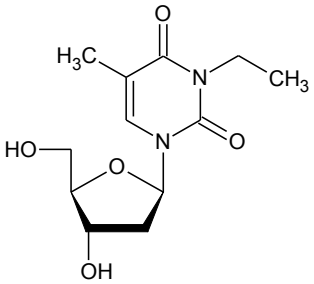
*N*3-ethyladenine



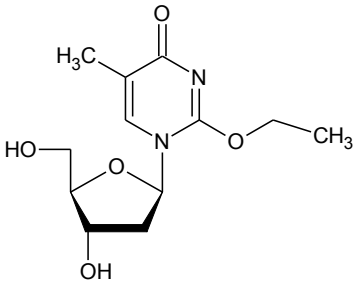
*N*7-ethylguanine



O^4-ethylthymidine



*N*3-ethylthymidine



O^2-ethylthymidine

Formation of Alkylated DNA from Endogenous Lipid Peroxidation and Glucose Metabolism

Unlike DNA adducts produced as a result from exposure to tobacco smoke, endogenous exposure to reactive species resulting from normal cellular processes is unavoidable. The oxidation of lipids, unsaturated fatty acids and carbohydrates results in the production of glyoxal and methylglyoxal as byproducts.[32-35] Methylglyoxal can also be generated from the degradation of glucose and is present at higher levels in the plasma, aortic cells, and urine of diabetic patients.[36-39] Both glyoxal and methylglyoxal are capable of interacting with DNA, causing damage.

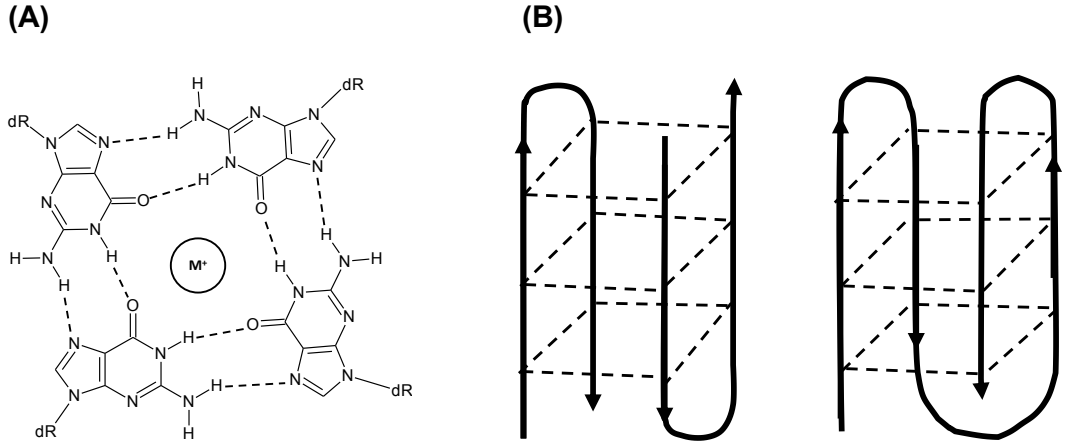
Under physiological conditions, N^2-(1-carboxyethyl)-2'-deoxyguanosine (N^2-CEdG) was the major stable DNA adduct formed when calf thymus DNA was exposed to methylglyoxal.[40] In addition, N^2-CEdG was detected in untreated human cells, and a dose-dependent increase in the lesion was observed for cells exposed to both glucose or methylglyoxal.[41] Aside from the detection of this lesion in human cells, in *in vitro* and *in vivo* studies of bacterial translesion synthesis (TLS) polymerases, *E. coli* pol IV, a DinB DNA polymerase was able to bypass N^2-CEdG in an error-free fashion.[41] Together, these studies support the role of N^2-CEdG as an endogenously produced DNA lesion which can be accurately and efficiently bypassed by DinB DNA polymerase, a DNA polymerase conserved in all three kingdoms of life. The conservation of the DinB polymerases suggests that certain DNA polymerases may have evolved to bypass specific DNA lesions encountered throughout the lifetime of the organism.

Formation of Non-B Structured DNA

In addition to damaged DNA bases which have been covalently modified, the DNA strand itself is capable of forming unusual secondary structures that can disrupt cellular functions. Under certain conditions, such as replication and transcription, specific single-stranded DNA sequences can adopt non-duplex (non-B) structures, including triplexes, cruciforms, slipped structures, Z-DNA and G-quadruplexes (G4).[42] These non-B DNA conformations have been implicated in genetic disorders and psychiatric diseases.[43, 44] G4 structures are particularly interesting due to their unique physical properties and, as revealed by computational studies, G4 sequences are ubiquitous in the genomes of both prokaryotes and eukaryotes.[45-47]

Single-stranded guanine-rich DNA and RNA sequences have been shown to fold into G4 structures,[48-51] consisting of two or more stacked guanine tetrads, *i.e.*, four guanine bases, stabilized through Hoogsteen hydrogen-bonding and electrostatic interactions with monovalent cations (Figure 1.2). It is notable that G4 structures have melting temperatures that are 20-30°C higher than to the melting temperatures of the corresponding duplex DNA.[51-53] These potential G4 sequences are conserved in related species[47, 54, 55] and are not randomly located throughout genomes, but instead are concentrated in the telomeres of eukaryotes, in micro- and minisatellite regions, in the promoter regions of genes, within ribosomal DNA, near transcription factor binding sites and at mitotic and meiotic double-strand break sites.[46, 47, 54-56] This non-random concentration of potential G4 structures in the genome suggests likely roles for these non-canonical structures in DNA replication, gene regulation and chromosome stability.

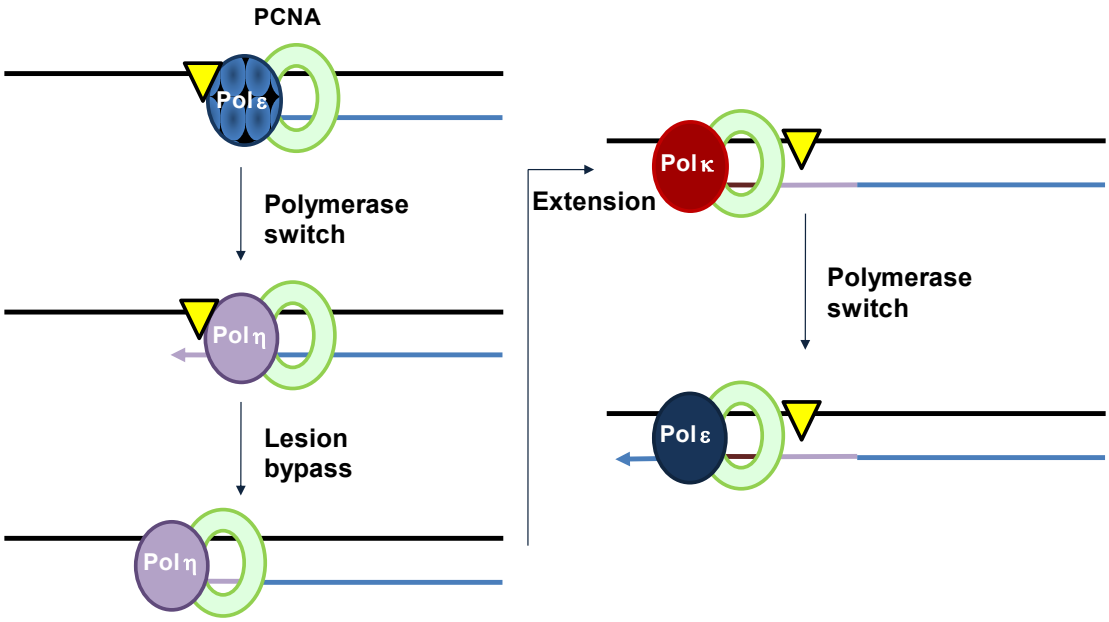
Figure 1.2. Structure of G-tetrad stabilized through Hoogsteen hydrogen bonding and electrostatic interactions with the monovalent cation (A) and intramolecular parallel and anti-parallel G-quadruplex structures (B).



Translesion Synthesis DNA Polymerases

To cope with DNA damage, cells are equipped with a variety of cellular mechanisms capable of sensing, removing, repairing and tolerating DNA damage. Due to their small active sites, most replicative DNA polymerases are unable to accommodate DNA lesions or resolve aberrant DNA structures on their own,[57, 58] resulting in stalled replication forks. As can be seen in Scheme 1.2, cells utilize specialized DNA polymerases called translesion synthesis (TLS) DNA polymerases as a mechanism to tolerate blocked DNA synthesis. Sensing a stalled replication fork, a TLS polymerase(s) (i.e., pol η in Scheme 1.2) is recruited to the primer terminus and switches out with the replicative polymerase (i.e. - pol ε in Scheme 1.2). The TLS polymerase then resumes replication, bypassing the DNA damage and inserting a nucleotide opposite the DNA damage; this bypass can occur in an error-free or error-prone fashion. Following insertion opposite the damage, the primer terminus is extended. Extension past the lesion is completed by the same TLS polymerase, or depending upon the DNA damage, a second TLS polymerase (i.e., pol κ in Scheme 1.2) may switch in and complete the extension.[59] Following synthesis of a short segment of DNA, the TLS polymerase then dissociates from the primer terminus and the replicative polymerase (i.e., pol ε in Scheme 1.2) resumes DNA synthesis. In this way, cells are able to tolerate DNA damage, escaping cellular checkpoints which may be activated following a stalled/collapsed replication fork.[57, 58]

Scheme 1.2. A polymerase switching model for translesion DNA synthesis.



Compared to their replicative counterparts, TLS polymerases are less processive and more error-prone when replicating undamaged DNA. This is due to the extra little finger domain that TLS polymerases possess.[60] This little finger domain results in a larger active site, which can accommodate bulkier DNA lesions and allow for the insertion of nucleotides opposite damaged bases.[57, 58, 61] In addition, TLS polymerases lack the proofreading 3'→5' exonuclease activity which is capable of removing mispaired nucleotides from the primer terminus.[57, 58, 61, 62]

Many of these TLS polymerases are conserved throughout all kingdoms of life and have, in some cases, been shown to efficiently and accurately bypass specific DNA lesions. For example, pol η is able to efficiently incorporate the correct nucleotide opposite UV induced *cis-syn* cyclobutane pyrimidine dimers.[63-66] The importance of pol η is highlighted by the genetic disease, the variant form of xeroderma pigmentosum. Individuals suffering from this disease carry an inactive form of pol η and are pre-disposed to sunlight-induced skin cancer.[63, 67] Other TLS polymerases, such as *Escherichia coli* (*E. coli*) pol IV and its human homologue pol κ, have been shown to accurately and efficiently bypass minor-groove N^2-alkyl guanine lesions both *in vitro* and *in vivo*.[41, 68, 69]

Steady-state Kinetic Assays

DNA polymerases, and in particular TLS DNA polymerases, have the potential to be highly mutagenic when processing damaged DNA bases. To study the potential mutagenicity of DNA polymerases, steady-state kinetic assays have been developed to

measure polymerase fidelity and efficiency to incorporate nucleotides.[70-72] In these assays, a DNA polymerase along with individual nucleotides are added to separate reaction mixtures containing a radiolabeled primer annealed to a complementary template containing a site-specifically incorporated damaged DNA base. Following termination of the reaction and separation of the products by polyacrylamide gel electrophoresis (PAGE), the rate of insertion for each nucleotide can be determined and the kinetic parameters of the DNA polymerase calculated.[70-72] While useful information about the activity of the DNA polymerase is obtained, steady-state kinetic assays have several drawbacks. In order to determine the fidelity and efficiency of a DNA polymerase, individual nucleotides are added to the reaction mixture, whereas *in vivo* all four nucleotides are available to the DNA polymerase. Unlike under cellular conditions, this results in no competition between the nucleotides during insertion. Another drawback to the steady-state kinetic assay is its inability to assess whether bypass of a damaged DNA base would result in an insertion or deletion during DNA synthesis. For exceptionally bulky nucleobase lesions, the likelihood of an insertion or deletion becomes much higher. Therefore other *in vitro* and *in vivo* analytical techniques are used to study the effect that damaged DNA bases have on replication.

Primer Extension Assays

Primer extension assays are typically performed to assess the relative activity that DNA polymerases have when replicating undamaged and damaged DNA templates. Unlike steady-state kinetic assays, in these experiments, all four nucleotides plus the

DNA polymerase are incubated with the radio-labeled primer annealed to the template DNA. Conventionally, the replication products are separated by PAGE, but products are separated based only upon size, and while the images generated by PAGE analysis are easy to interpret and quantify, determining sequence information can be challenging. The development of soft ionization techniques such as electrospray ionization (ESI), which couples liquid chromatography and mass spectrometry, allows for the facile sequence determination of replication products generated during primer extension reactions. Several studies investigating the effect that DNA damage has on replication have been conducted using this technique.[73-76]

During ESI, multiply charged ions are generated.[77] For DNA sequences longer than twenty bases, resolving and determining the sequence composition of oligodeoxynucleotides (ODNs) become very challenging. For primer extension reactions, however, the primer/template complex must be sufficiently long for the DNA polymerase to have enough points of contact to interact with the duplex DNA.

Initially, groups developing primer extension LC-MS/MS assays incorporated a uracil base in the primer located several bases from the 3' end of the primer, and following completion of the replication reaction, the mixture is treated with uracil DNA glycosylase (UDG) which creates an abasic site. This abasic site is then cleaved by treating the reaction mixture with piperidine at 95°C.[73-76] In this way, the extension products produced during the replication reaction are shortened to smaller ODN sequences that are effectively analyzed by LCMS. The reaction conditions necessary to remove the uracil base and cleave the abasic site are very harsh, however, and may cause

decomposition of the lesion in the damaged template, introducing uncertainty whether the replication products were synthesized from a damaged or undamaged template. One possible alternative to the uracil/piperidine approach is to take advantage of the DNA clipping activity restriction enzymes. These enzymes are capable of recognizing and excising specific DNA sequences under similar reaction conditions as those for the replication reactions. Therefore, by engineering a restriction site into the primer/template, following termination of the replication reaction, the resulting mixture can be treated with a restriction enzyme to generate short ODN fragments amenable to analysis via LC-MS/MS.

Competitive Replication and Adduct Bypass Assay

While steady-state and primer extension assays offer insights into the role that individual polymerases play when replicating past damaged DNA bases, these assays are performed *in vitro* with purified DNA polymerases. *In vivo* cellular replication involves the coordinated action of protein complexes. In particular, DNA polymerases synthesize new strands of DNA by working in concert with a host of accessory proteins. To more accurately study the effect damaged nucleobases have on DNA replication, *in vivo* replication assays are capable assessing the true effect damaged nucleobases may exert on DNA replication. Delaney and Essigmann[78] have developed a competitive replication and adduct bypass (CRAB) assay to interrogate the effect that a site-specifically inserted DNA lesion has on DNA replication. This CRAB assay takes advantage of a self-replicating single-stranded bacteriophage which can be transformed into *E. coli* host

cells. Single-stranded stranded genomes containing the site-specifically incorporated DNA damaged along with control and competitor genomes are constructed via enzymatic ligation.[78] Following transformation into *E. coli* cells, the ratio of genomes arising from replication of the damage-containing and control genomes can be normalized to that of the competitor genome. This allows for the determination of bypass efficiency of replication machinery with respect to the incorporated DNA damage. If a lesion is particularly blocking to replicative polymerases, by transforming cells into TLS polymerase-deficient *E. coli* cells, the polymerase involved in bypassing the lesion can be determined. One important aspect of the CRAB assay is that ssDNA is generally not subject to DNA repair and thus the progeny genomes are the result of replication opposite the inserted DNA damage.[79]

To date, these methods have primarily been used to investigate damaged nucleobases, however these same methods can be applied to non-B DNA structures present in cells. In this regard, the effect that uniquely structured DNA, i.e. G4 structure, has on DNA replication can also be investigated.

Scope of this dissertation

In the second two chapters of this dissertation, the effect alkylated thymidines have on DNA replication is investigated. In Chapter 2, two lesions, O^2- and O^4-methylthymidine (O^2-MdT and O^4-MdT) are studied, while in Chapter 3 a systematic study of the effects that O^2-EtdT, $N3$-EtdT, and O^4-EtdT have on DNA replication is explored. The role that eukaryotic TLS polymerases play in processing endogenously

produced N^2-CEdG DNA lesions is presented in Chapter 4. Finally, in Chapter 5, the effect of two G4 sequences, one derived from the *c-Myb* proto-oncogene and the other from the *c-Kit* oncogene, have on DNA replication is explored.

References

(1) Finkel, T., and Holbrook, N. J. (2000) Oxidants, oxidative stress and the biology of ageing. *Nature*, *408*, 239-247.

(2) Heck, D. E., Vetrano, A. M., Mariano, T. M., and Laskin, J. D. (2003) UVB light stimulates production of reactive oxygen species: unexpected role for catalase. *J. Biol. Chem.*, *278*, 22432-22436.

(3) Cadet, J., Sage, E., and Douki, T. (2005) Ultraviolet radiation-mediated damage to cellular DNA. *Mutat. Res.*, *571*, 3-17.

(4) Bellon, S., Ravanat, J. L., Gasparutto, D., and Cadet, J. (2002) Cross-linked thymine-purine base tandem lesions: synthesis, characterization, and measurement in γ-irradiated isolated DNA. *Chem. Res. Toxicol.*, *15*, 598-606.

(5) Friedman, M., and Rasooly, R. (2013) Review of the inhibition of biological activities of food-related selected toxins by natural compounds. *Toxins (Basel)*, *5*, 743-775.

(6) Prevost, V., and Shuker, D. E. (1996) Cigarette smoking and urinary 3-alkyladenine excretion in man. *Chem. Res. Toxicol.*, *9*, 439-444.

(7) Marnett, L. J., and Plastaras, J. P. (2001) Endogenous DNA damage and mutation. *Trends Genet.*, *17*, 214-221.

(8) Friedberg, E. C., McDaniel, L. D., and Schultz, R. A. (2004) The role of endogenous and exogenous DNA damage and mutagenesis. *Curr. Opin. Genet. Dev.*, *14*, 5-10.

(9) Giovino, G. A., Mirza, S. A., Samet, J. M., Gupta, P. C., Jarvis, M. J., Bhala, N., Peto, R., Zatonski, W., Hsia, J., Morton, J., Palipudi, K. M., and Asma, S. (2012) Tobacco use in 3 billion individuals from 16 countries: an analysis of nationally representative cross-sectional household surveys. *Lancet*, *380*, 668-679.

(10) Dietz, N. A., Lee, D. J., Fleming, L. E., Leblanc, W. G., McCollister, K. E., Arheart, K. L., Davila, E. P., and Caban-Martinez, A. J. (2011) Trends in smokeless tobacco use in the us workforce: 1987-2005. *Tob Induc Dis*, *9*, 6.

(11) (2004) Tobacco smoke and involuntary smoking. *IARC Monographs Evaluation of Carcinogen Risk to Humans*, *83*, 1-1438.

(12) Cancer., I. A. f. R. o. Tobacco smoke and involuntary smoking IARC monographs on the evaluation of carcinogenic risks to humans, pp 1179-1187, IARC ; Distributed by IARC Press and the World Health Organization Distribution and Sales, Lyon, France; Geneva.

(13) Cancer, I. A. f. R. o., Ed. (2007) *Smokeless tobacco and tobacco-specific nitrosamines*, Vol. 89, IARC, Lyon, France.

(14) Peterson, L. A. (2010) Formation, repair, and genotoxic properties of bulky DNA adducts formed from tobacco-specific nitrosamines. *J. Nucleic Acids*, *2010*.

(15) Hecht, S. S., Spratt, T. E., and Trushin, N. (1988) Evidence for 4-(3-pyridyl)-4-oxobutylation of DNA in F344 rats treated with the tobacco-specific nitrosamines 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone and *N*'-nitrosonornicotine. *Carcinogenesis*, *9*, 161-165.

(16) Shrivastav, N., Li, D., and Essigmann, J. M. (2010) Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis*, *31*, 59-70.
(17) Wang, L., Spratt, T. E., Liu, X. K., Hecht, S. S., Pegg, A. E., and Peterson, L. A. (1997) Pyridyloxobutyl adduct O^6-[4-oxo-4-(3-pyridyl)butyl]guanine is present in 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone-treated DNA and is a substrate for O^6-alkylguanine-DNA alkyltransferase. *Chem. Res. Toxicol.*, *10*, 562-567.
(18) Wang, M., Cheng, G., Sturla, S. J., Shi, Y., McIntee, E. J., Villalta, P. W., Upadhyaya, P., and Hecht, S. S. (2003) Identification of adducts formed by pyridyloxobutylation of deoxyguanosine and DNA by 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone, a chemically activated form of tobacco specific carcinogens. *Chem. Res. Toxicol.*, *16*, 616-626.
(19) Upadhyaya, P., Sturla, S. J., Tretyakova, N., Ziegel, R., Villalta, P. W., Wang, M., and Hecht, S. S. (2003) Identification of adducts produced by the reaction of 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanol with deoxyguanosine and DNA. *Chem. Res. Toxicol.*, *16*, 180-190.
(20) Hecht, S. S., Villalta, P. W., Sturla, S. J., Cheng, G., Yu, N., Upadhyaya, P., and Wang, M. (2004) Identification of O^2-substituted pyrimidine adducts formed in reactions of 4-(acetoxymethylnitrosamino)- 1-(3-pyridyl)-1-butanone and 4-(acetoxymethylnitros- amino)-1-(3-pyridyl)-1-butanol with DNA. *Chem. Res. Toxicol.*, *17*, 588-597.
(21) Upadhyaya, P., McIntee, E. J., Villalta, P. W., and Hecht, S. S. (2006) Identification of adducts formed in the reaction of 5'-acetoxy-*N*'-nitrosonornicotine with deoxyguanosine and DNA. *Chem. Res. Toxicol.*, *19*, 426-435.
(22) Singh, R., Kaur, B., and Farmer, P. B. (2005) Detection of DNA damage derived from a direct acting ethylating agent present in cigarette smoke by use of liquid chromatography-tandem mass spectrometry. *Chem. Res. Toxicol.*, *18*, 249-256.
(23) Chao, M. R., Wang, C. J., Chang, L. W., and Hu, C. W. (2006) Quantitative determination of urinary N7-ethylguanine in smokers and non-smokers using an isotope dilution liquid chromatography/tandem mass spectrometry with on-line analyte enrichment. *Carcinogenesis*, *27*, 146-151.
(24) Anna, L., Kovacs, K., Gyorffy, E., Schoket, B., and Nair, J. (2011) Smoking-related O^4-ethylthymidine formation in human lung tissue and comparisons with bulky DNA adducts. *Mutagenesis*, *26*, 523-527.
(25) Bhanot, O. S., Grevatt, P. C., Donahue, J. M., Gabrielides, C. N., and Solomon, J. J. (1992) *In vitro* DNA replication implicates O^2-ethyldeoxythymidine in transversion mutagenesis by ethylating agents. *Nucleic Acids Res.*, *20*, 587-594.
(26) Balbo, S., Villalta, P. W., and Hecht, S. S. (2011) Quantitation of 7-ethylguanine in leukocyte DNA from smokers and nonsmokers by liquid chromatography-nanoelectrospray-high resolution tandem mass spectrometry. *Chem. Res. Toxicol.*, *24*, 1729-1734.

(27) Chen, H. J., Wang, Y. C., and Lin, W. P. (2012) Analysis of ethylated thymidine adducts in human leukocyte DNA by stable isotope dilution nanoflow liquid chromatography-nanospray ionization tandem mass spectrometry. *Anal. Chem.*, *84*, 2521-2527.

(28) Pauly, G. T., Peterson, L. A., and Moschel, R. C. (2002) Mutagenesis by O^6-[4-oxo-4-(3-pyridyl)butyl]guanine in *Escherichia coli* and human cells. *Chem. Res. Toxicol.*, *15*, 165-169.

(29) Wang, M., Cheng, G., Villalta, P. W., and Hecht, S. S. (2007) Development of liquid chromatography electrospray ionization tandem mass spectrometry methods for analysis of DNA adducts of formaldehyde and their application to rats treated with *N*-nitrosodimethylamine or 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Chem. Res. Toxicol.*, *20*, 1141-1148.

(30) Lao, Y., Yu, N., Kassie, F., Villalta, P. W., and Hecht, S. S. (2007) Formation and accumulation of pyridyloxobutyl DNA adducts in F344 rats chronically treated with 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone and enantiomers of its metabolite, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol. *Chem. Res. Toxicol.*, *20*, 235-245.

(31) Li, L., Perdigao, J., Pegg, A. E., Lao, Y., Hecht, S. S., Lindgren, B. R., Reardon, J. T., Sancar, A., Wattenberg, E. V., and Peterson, L. A. (2009) The influence of repair pathways on the cytotoxicity and mutagenicity induced by the pyridyloxobutylation pathway of tobacco-specific nitrosamines. *Chem. Res. Toxicol.*, *22*, 1464-1472.

(32) Frankel, E. N. (1983) Volatile lipid oxidation products. *Prog. Lipid Res.*, *22*, 1-33.

(33) Loidl-Stahlhofen, A., and Spiteller, G. (1994) α-Hydroxyaldehydes, products of lipid peroxidation. *Biochim. Biophys. Acta.*, *1211*, 156-160.

(34) Mizutari, K., Ono, T., Ikeda, K., Kayashima, K., and Horiuchi, S. (1997) Photo-enhanced modification of human skin elastin in actinic elastosis by N^{ε}-carboxymethyl-lysine, one of the glycoxidation products of the Maillard reaction. *J. Invest. Dermatol.*, *108*, 797-802.

(35) Thornalley, P. J., Langborg, A., and Minhas, H. S. (1999) Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J.*, *344 Pt 1*, 109-116.

(36) Lapolla, A., Flamini, R., Dalla Vedova, A., Senesi, A., Reitano, R., Fedele, D., Basso, E., Seraglia, R., and Traldi, P. (2003) Glyoxal and methylglyoxal levels in diabetic patients: quantitative determination by a new GC/MS method. *Clin. Chem. Lab. Med.*, *41*, 1166-1173.

(37) Raj, D. S., Choudhury, D., Welbourne, T. C., and Levi, M. (2000) Advanced glycation end products: a Nephrologist's perspective. *Am. J. Kidney Dis.*, *35*, 365-380.

(38) Schneider, M., Thoss, G., Hubner-Parajsz, C., Kientsch-Engel, R., Stahl, P., and Pischetsrieder, M. (2004) Determination of glycated nucleobases in human urine by a new monoclonal antibody specific for N^2-carboxyethyl-2'-deoxyguanosine. *Chem. Res. Toxicol.*, *17*, 1385-1390.

(39) Li, H., Nakamura, S., Miyazaki, S., Morita, T., Suzuki, M., Pischetsrieder, M., and Niwa, T. (2006) N^2-carboxyethyl-2'-deoxyguanosine, a DNA glycation marker, in kidneys and aortas of diabetic and uremic patients. *Kidney Int.*, *69*, 388-392.

(40) Frischmann, M., Bidmon, C., Angerer, J., and Pischetsrieder, M. (2005) Identification of DNA adducts of methylglyoxal. *Chem. Res. Toxicol.*, *18*, 1586-1592.

(41) Yuan, B., Cao, H., Jiang, Y., Hong, H., and Wang, Y. (2008) Efficient and accurate bypass of N^2-(1-carboxyethyl)-2'-deoxyguanosine by DinB DNA polymerase *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. U S A*, *105*, 8679-8684.

(42) Wang, G., and Vasquez, K. M. (2006) Non-B DNA structure-induced genetic instability. *Mutat. Res.*, *598*, 103-119.

(43) Bacolla, A., and Wells, R. D. (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol. Carcinog.*, *48*, 273-285.

(44) Zhao, J., Bacolla, A., Wang, G., and Vasquez, K. M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell Mol. Life Sci.*, *67*, 43-62.

(45) Todd, A. K., Johnston, M., and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, *33*, 2901-2907.

(46) Hershman, S. G., Chen, Q., Lee, J. Y., Kozak, M. L., Yue, P., Wang, L. S., and Johnson, F. B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, *36*, 144-156.

(47) Rawal, P., Kummarasetti, V. B., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S. K., and Chowdhury, S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.*, *16*, 644-655.

(48) Sen, D., and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, *334*, 364-366.

(49) Gilbert, D. E., and Feigon, J. (1999) Multistranded DNA structures. *Curr. Opin. Struct. Biol.*, *9*, 305-314.

(50) Neidle, S., and Parkinson, G. N. (2003) The structure of telomeric DNA. *Curr. Opin. Struct. Biol.*, *13*, 275-283.

(51) Lane, A. N., Chaires, J. B., Gray, R. D., and Trent, J. O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, *36*, 5482-5515.

(52) Bochman, M. L., Paeschke, K., and Zakian, V. A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, *13*, 770-780.

(53) Lipps, H. J., and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, *19*, 414-422.

(54) Capra, J. A., Paeschke, K., Singh, M., and Zakian, V. A. (2010) G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput. Biol.*, *6*, e1000861.

(55) Nakken, S., Rognes, T., and Hovig, E. (2009) The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. *Nucleic Acids Res.*, *37*, 5749-5756.
(56) Huppert, J. L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.*, *37*, 1375-1384.
(57) Lehmann, A. R., Niimi, A., Ogi, T., Brown, S., Sabbioneda, S., Wing, J. F., Kannouche, P. L., and Green, C. M. (2007) Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst)*, *6*, 891-899.
(58) Sale, J. E., Lehmann, A. R., and Woodgate, R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell. Biol.*, *13*, 141-152.
(59) Bridges, B. A., and Woodgate, R. (1985) The two-step model of bacterial UV mutagenesis. *Mutat. Res.*, *150*, 133-139.
(60) Lange, S. S., Takata, K., and Wood, R. D. (2011) DNA polymerases and cancer. *Nat. Rev. Cancer*, *11*, 96-110.
(61) Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G. C., Wang, Z., and Woodgate, R. (2001) The Y-family of DNA polymerases. *Mol. Cell*, *8*, 7-8.
(62) Nohmi, T. (2006) Environmental stress and lesion-bypass DNA polymerases. *Annu. Rev. Microbiol.*, *60*, 231-253.
(63) Johnson, R. E., Prakash, S., and Prakash, L. (1999) Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Pol η. *Science*, *283*, 1001-1004.
(64) Johnson, R. E., Washington, M. T., Prakash, S., and Prakash, L. (2000) Fidelity of human DNA polymerase η. *J. Biol. Chem.*, *275*, 7447-7450.
(65) Washington, M. T., Johnson, R. E., Prakash, S., and Prakash, L. (2000) Accuracy of thymine-thymine dimer bypass by Saccharomyces cerevisiae DNA polymerase η. *Proc. Natl. Acad. Sci. U S A*, *97*, 3094-3099.
(66) Yoon, J. H., Prakash, L., and Prakash, S. (2009) Highly error-free role of DNA polymerase η in the replicative bypass of UV-induced pyrimidine dimers in mouse and human cells. *Proc. Natl. Acad. Sci. U S A*, *106*, 18219-18224.
(67) Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature*, *399*, 700-704.
(68) Choi, J. Y., Angel, K. C., and Guengerich, F. P. (2006) Translesion synthesis across bulky N^2-alkyl guanine DNA adducts by human DNA polymerase κ. *J. Biol. Chem.*, *281*, 21062-21072.
(69) Yuan, B., You, C., Andersen, N., Jiang, Y., Moriya, M., O'Connor, T. R., and Wang, Y. The roles of DNA polymerases κ and ι in the error-free bypass of N^2-carboxyalkyl-dG lesions in mammalian cells. *J. Biol. Chem.*
(70) Fersht, A. (1985) *Enzyme Structure and Mechanism. 2.*, W.H. Freeman & Co., New York.

(71) Bertram, J. G., Oertell, K., Petruska, J., and Goodman, M. F. (2010) DNA polymerase fidelity: comparing direct competition of right and wrong dNTP substrates with steady state and pre-steady state kinetics. *Biochemistry*, *49*, 20-28.
(72) Creighton, S., and Goodman, M. F. (1995) Gel kinetic analysis of DNA polymerase fidelity in the presence of proofreading using bacteriophage T4 DNA polymerase. *J. Biol. Chem.*, *270*, 4759-4774.
(73) Chowdhury, G., and Guengerich, F. P. (2011) Liquid chromatography-mass spectrometry analysis of DNA polymerase reaction products. *Curr. Protoc. Nucleic. Acid Chem.*, *Chapter 7*, Unit 7 16 11-11.
(74) Gu, C., and Wang, Y. (2007) *In vitro* replication and thermodynamic studies of methylation and oxidation modifications of 6-thioguanine. *Nucleic Acids Res.*, *35*, 3693-3704.
(75) Cho, S. H., and Guengerich, F. P. (2013) Replication past the butadiene diepoxide-derived DNA adduct *S*-[4-(*N*^{6}-deoxyadenosinyl)-2,3-dihydroxybutyl]glutathione by DNA polymerases. *Chem. Res. Toxicol.*, *26*, 1005-1013.
(76) Kotapati, S., Maddukuri, L., Wickramaratne, S., Seneviratne, U., Goggin, M., Pence, M. G., Villalta, P., Guengerich, F. P., Marnett, L., and Tretyakova, N. (2012) Translesion synthesis across 1,*N*^{6}-(2-hydroxy-3-hydroxymethylpropan-1,3-diyl)-2'-deoxyadenosine (1,*N*^{6}-γ-HMHP-dA) adducts by human and archebacterial DNA polymerases. *J. Biol. Chem.*, *287*, 38800-38811.
(77) Ho, C. S., Lam, C. W., Chan, M. H., Cheung, R. C., Law, L. K., Lit, L. C., Ng, K. F., Suen, M. W., and Tai, H. L. (2003) Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin. Biochem. Rev.*, *24*, 3-12.
(78) Delaney, J. C., and Essigmann, J. M. (2006) Assays for determining lesion bypass efficiency and mutagenicity of site-specific DNA lesions *in vivo*. *Methods Enzymol.*, *408*, 1-15.
(79) Moriya, M. (1993) Single-stranded shuttle phagemid for mutagenesis studies in mammalian cells: 8-oxoguanine in DNA induces targeted G.C-->T.A transversions in simian kidney cells. *Proc. Natl. Acad. Sci. U. S. A.*, *90*, 1122-1126.

Chapter 2: *In-vitro* Replication Studies on O^2-Methylthymidine and O^4-Methylthymidine

Introduction

Low levels of *N*-nitroso compounds (NOCs) are present in a variety of sources including air, beer, food, water, chewing tobacco and tobacco smoke. Many of these NOCs are known carcinogens in both humans and laboratory animals and, after metabolic activation, they can result in DNA alkylation.[1] Alkylation at nitrogen atoms on nucleobases has been shown to induce transversions, frameshift mutations, and small deletions.[2] Alkylation at oxygen atoms, albeit occurring to a lesser extent than *N*-alkylation,[3, 4] primarily produces point mutations and is linked with the carcinogenic and mutagenic properties of many alkylating agents.[5] For larger alkylating agents, the reactivity of oxygen atoms in nucleobases differs, with the O^6 position of guanine being the most reactive, followed by the O^2 and O^4 positions of thymine, and then the O^2 position of cytosine.[3, 6]

Different *O*-alkylated DNA lesions can be repaired with varying efficiencies. For instance, O^6-alkylguanine lesions are readily repaired, whereas the *O*-alkylpyrimidines are removed at much lower rates. In this vein, *Escherichia coli* O^6-methylguanine-DNA methyltransferase (Ada) and mammalian O^6-methylguanine-DNA methyltransferase (MGMT) repair O^6-MdG 10^5 and ~10^3 times faster than O^4-methylthymidine (O^4-MdT), respectively.[7-10] Various alkylated purine and pyrimidine bases, including O^2- and O^4-alkylthymines, have been found in tissues of rats treated with *N,N*-dimethylnitrosamine

(DMN) or *N*-ethyl-*N*-nitrosourea (ENU).[11] Interestingly the amount of O^4-ethylthymine in tissues was lower than that of O^2-ethylthymine, suggesting that the latter might be more resistant to repair.[11] The poor repair of *O*-alkylthymidines may render these lesions highly persistent in the genome, thereby interfering with the cellular replication machinery.

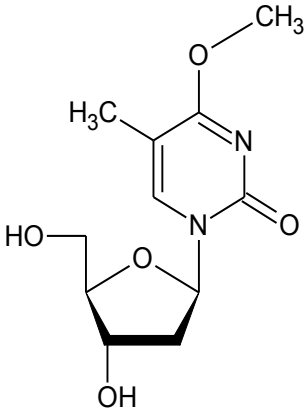
To cope with unrepaired DNA lesions which stall DNA replication, cells are equipped with a number of translesion synthesis (TLS) DNA polymerases capable of replicating past damaged nucleobases.[12, 13] These specialized polymerases have lower fidelity and processivity than replicative DNA polymerases, largely owing to their more spacious active sites and lack of proofreading 3'→5' exonuclease activities.[14] However, in some instances DNA synthesis mediated by TLS polymerases is both accurate and efficient. For example, the DinB DNA polymerase (i.e., pol κ in mammalian cells and pol IV in *Escherichia coli*), a Y-family polymerase conserved in all three kingdoms of life, is capable of bypassing accurately and efficiently some minor-groove N^2-substituted dG derivatives *in vitro* and in cells.[15-18] Polymerase η, another Y-family DNA polymerase, incorporates, with high efficiency, the correct nucleotide opposite the *cis-syn* cyclobutane pyrimidine dimers induced by UV light,[19-21] The importance of TLS DNA polymerases is manifested by the fact that mutational inactivation of pol η in humans leads to the variant form of xeroderma pigmentosum,[22, 23] a genetic disease characterized by an elevated predisposition for sunlight-induced skin cancers.[24]

In addition to their poor repair, several studies have shown that O^4- and O^2-alkylthymidine are both blocking and miscoding during DNA replication *in vitro* and *in*

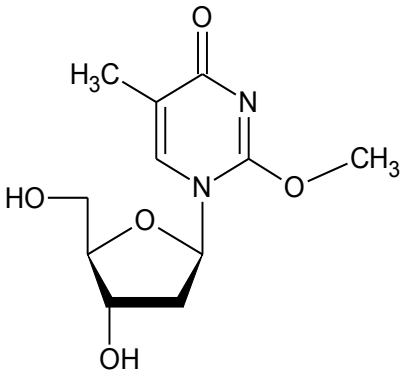
vivo. O^4-alkylthymidines have been shown to primarily induce T→C mutations.[2, 6, 10, 25-29] The minor-groove lesion, O^2-ethylthymidine was found to be highly blocking to the exonuclease-free Klenow fragment of *E. coli* DNA polymerase I (Kf⁻) and the lesion directs the polymerase to incorporate both dA and dT.[30] In addition, O^2-methylthymidine (O^2-MdT) and O^2-[4-(3-pyridyl)-4-oxobut-1-yl]thymidine (O^2-POB-dT) are strongly blocking to replication mediated by Kf⁻ and Dpo4 *in vitro* and the replication machinery of *E. coli* cells.[31-34] Moreover, both lesions exhibit strong miscoding potential in these experimental systems. [31, 32]

Herein we investigated how O^2-MdT and O^4-MdT perturb DNA replication by three purified polymerases *in vitro* (Methylated thymidine structures shown in Figure 2.1). These included the Kf⁻ and two Y-family polymerases, *Saccharomyces cerevisiae* DNA polymerase η (pol η) and human DNA polymerase κ. We chose these polymerases because Kf⁻ is a widely used model DNA polymerase for examining how DNA lesions perturb DNA replication, pol η is capable of bypassing another major-groove *O*-alkyl product, O^6-methylguanine,[35] and pol κ can accurately and efficiently bypass a number of minor-groove N^2-dG lesions.[15, 17, 18]

Figure 2.1. Structure of *O*-methylated thymidines used in this study.



O[4]-methylthymidine



O[2]-methylthymidine

Materials and Methods

Materials

All enzymes and chemicals unless otherwise specified were purchased from New England Biolabs (Ipswich, MA) or Sigma-Aldrich (St. Louis, MO). Unmodified oligodeoxyribonucleotides (ODNs) used in this study were purchased from Integrated DNA Technologies (Coralville, IA). [γ-^{32}P]ATP was obtained from Perkin Elmer (Piscataway, NJ). 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was purchased from TCI America (Portland, OR). The phosphoramidite building block of O^4-MdT, conventional phosphoramidites of unmodified nucleosides, and other reagents for solid-phase DNA synthesis were obtained from Glen Research (Sterling, VA). Kf$^-$ and human DNA polymerase κ were from New England Biolabs (Ipswich, MA) and Enzymax (Lexington, KY), respectively. *Saccharomyces cerevisiae* DNA polymerase η was expressed and purified following previously published procedures.[36, 37]

Preparation of Lesion-containing Substrates

The O^2-MdT phosphoramidite building block was synthesized following previously published procedures.[38] ODNs containing a site-specifically incorporated O^2-MdT or O^4-MdT [17mer, d(CCATGGCAXGAGAATTC), X = O^2-MdT or O^4-MdT] were synthesized on a Beckman Oligo 1000S DNA Synthesizer (Fullerton, CA). The ODNs were cleaved from the controlled pore glass (CPG) support, deprotected by treatment with 10% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in methanol in the dark at 23°C for five days and purified by 20% denaturing polyacrylamide gel electrophoresis (PAGE). The structures of the ODNs were confirmed by electrospray ionization-mass

spectrometry (ESI-MS) and tandem MS (MS/MS) analyses (Figure A.1). The 28mer substrates, d(CCATGGCAXGAGAATTCTATGGTCCTAG) ('X' = dT, O^2-MdT or O^4-MdT) were obtained by ligating the above-described 17-mer ODNs with a 5'-phosphorylated d(TATGGTCCTAG) in the presence of a template ODN following previously published procedures.[39] The desired lesion-containing 28mer ODNs were purified using PAGE and desalted by ethanol precipitation. The purity of the product was further confirmed by PAGE analysis.

Steady-state Kinetic Measurements – Nucleotide Insertion

In-vitro replication experiments were performed following previously described procedures.[40, 41] Briefly, the 28mer template, d(CCATGGCAXGAGAATTCTATGGTCCTAG) ('X' = dT, O^2-MdT or O^4-MdT, 20 nM), was annealed with a 5'-^{32}P-labeled 20-mer primer, d(GCTAGGACCATAGAATTCTC) (10 nM) for steady-state kinetic measurements of nucleotide insertion opposite the lesions and unmodified dT. For Kf$^-$, the reactions were carried out in a buffer containing 50 mM Tris-HCl (pH 7.5), 20 mM $MgCl_2$, 2 mM EDTA, 1.6 mM β-mercaptoethanol, and 5 μg/mL BSA. For pol η and pol κ, the reactions were conducted in a buffer containing 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, and 7.5 mM DTT.

For steady-state kinetic measurements monitoring the nucleotide insertion opposite the lesion, the primer-template complex was incubated with Kf$^-$ (0.7-7 nM), human pol κ (1-5 nM) or yeast pol η (7.5 nM) in the presence of an individual dNTP at varying concentrations. The reaction was continued at 23°C (pol κ) or 37°C (Kf$^-$ and pol

η) for 10 min in the same reaction buffers as described above, and terminated by adding a 2-volume excess of formamide gel-loading buffer. The buffer contained 80% formamide, 10 mM EDTA (pH 8.0), 1 mg/mL xylene cyanol and 1 mg/mL bromophenol blue. The products were resolved on 20% (29:1) cross-linked polyacrylamide gels containing 8 M urea. Gel band intensities for the substrates and products were quantified using a Typhoon 9410 variable-mode imager (Amersham Biosciences Co.) and ImageQuant version 5.2 (Amersham Biosciences Co.). The dNTP concentration was optimized for different insertion reactions to allow for approximately 20% primer extension.[42] The observed rate of dNTP incorporation (V_{obs}) was plotted versus the dNTP concentration, and the apparent steady-state kinetic parameters (K_m and V_{max}) for the incorporation of the correct and incorrect nucleotides were determined by fitting the rate data with the Michaelis-Menten equation:

$$V_{obs} = \frac{V_{max} \times [dNTP]}{K_m + [dNTP]}$$

The k_{cat} values were then calculated by dividing the V_{max} values with the concentration of the polymerase used. The efficiency of nucleotide incorporation was determined by the ratio of k_{cat}/K_m, and the fidelity of nucleotide incorporation was gauged by the frequency of misincorporation (f_{inc}), which was calculated using the following equation:[43]

$$f_{inc} = \frac{(k_{cat} / K_m)_{incorrect}}{(k_{cat} / K_m)_{correct}}$$

Steady-state Kinetic Measurements - Nucleotide Extension Mediated by Human pol κ

The dT-, O^2-MdT-, or O^4-MdT-containing templates (20 nM) were incubated with a 5'-^{32}P-21 mer primer d(GCTAGGACCATAGAATTCTCN) (10 nM), where N is an A

or G. The primer/template complex was incubated in the above-described buffer with pol κ (1-5 nM) and varying concentrations of the correct dTTP at 23°C for 10 min. The reaction was again terminated with formamide gel-loading buffer. The efficiency of extension was calculated by taking the ratio of k_{cat}/K_m, while the frequencies of mis-extension were determined by dividing the k_{cat}/K_m values for the extension with the incorrect nucleotide over that with the correct nucleotide.

Primer Extension Assays Monitored by Gel Electrophoresis

For primer extension assays, all four dNTPs and varying concentrations of a DNA polymerase, as indicated in the figures, were subsequently added to the duplex mixture and incubated for 60 min. The reaction was terminated by adding a 2-volume excess of formamide gel-loading buffer. The products were resolved on 20% (29:1) cross-linked polyacrylamide gels containing 8 M urea. Gel band intensities for the substrates and products were quantified by using a Typhoon 9410 variable-mode imager and ImageQuant version 5.2.

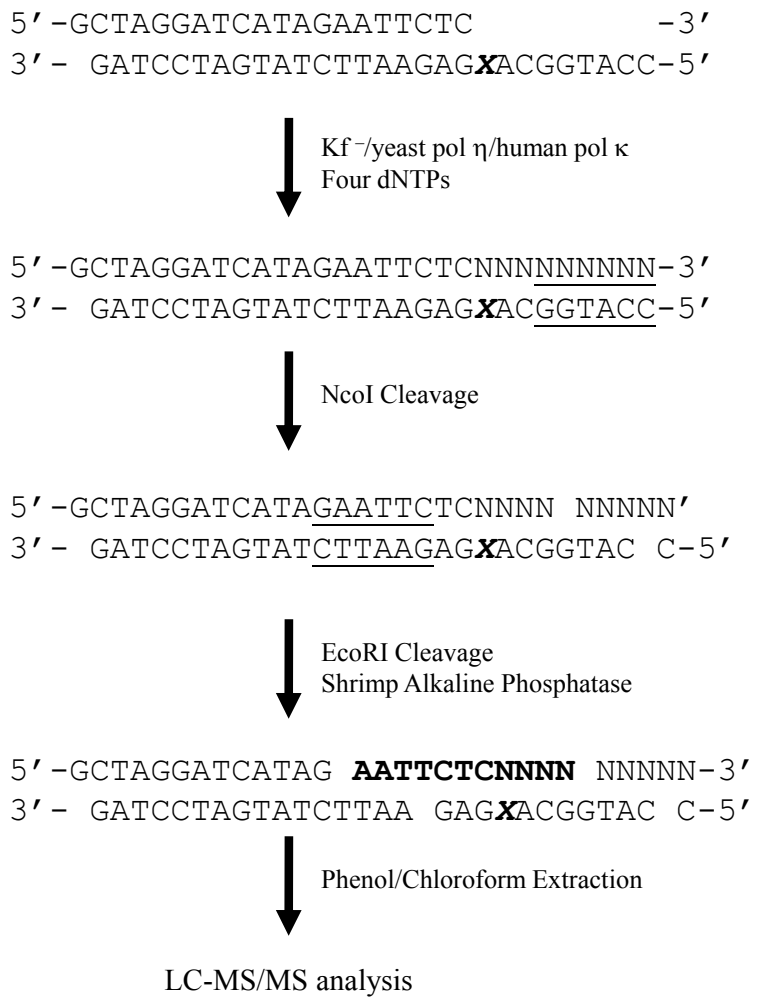
Primer Extension Assays Monitored by LC-MS/MS

The primer and lesion-containing templates (0.50 μM each) were annealed and incubated overnight in a 50-μL solution with pol κ (0.16 μM) at 23°C, and with pol η (0.24 μM) and Kf$^-$ (55 nM) at 37°C. The reactions were conducted in the presence of all four dNTPs (1 mM each) in the same buffers as described above. Two additional primer extension reactions were carried out for the O^4-MdT-containing substrate using pol η under the same conditions except that lower concentrations of dNTPs (50 μM each) or a shorter reaction time (6 hrs) was employed. Each replication reaction was subsequently

terminated by chloroform extraction and the aqueous layer dried using a Speed-Vac. The resulting replication products were cleaved sequentially with two restriction enzymes as shown in Scheme 2.1. In this regard, the replication mixture was first incubated with NcoI (40 U) in a buffer containing 50 mM NaCl, 10 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, and 1 mM DTT for 4 hrs. The NcoI was subsequently removed by chloroform extraction and the aqueous layer dried using a Speed-Vac. EcoRI (50 U) and shrimp alkaline phosphatase (20 U) were then added to the reaction mixture and incubated at 23°C for overnight in a buffer containing 50 mM NaCl, 10 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, and 1 mM DTT. The proteins in the mixture were again removed by chloroform extraction, and the aqueous layer dried using a Speed-Vac. The dried residue was reconstituted in 50-μL H_2O, and a 10-μL aliquot was injected for LC-MS/MS analysis on an Agilent Zorbax SB-C18 column (0.5 × 250 mm, 5 μm in particle size). The gradient for the HPLC elution was 5-min of 5-20% methanol followed by 35-min of 20-50% methanol in 400 mM HFIP (pH adjusted to 7.0). The temperature for the ion-transport tube was maintained at 350°C to minimize the formation of HFIP adducts of the ODNs.

To identify the replication products, samples were first analyzed in the data-dependent scan mode, where the most abundant ion found in MS was chosen for fragmentation in MS/MS. The fragment ions found in the MS/MS were manually assigned and the sequences of the ODNs determined. The mass spectrometer was subsequently set up for monitoring specifically the fragmentation of the precursor ions for the extended fragments of the primer strand.

Scheme 2.1. Restriction digestion method for replication reactions.



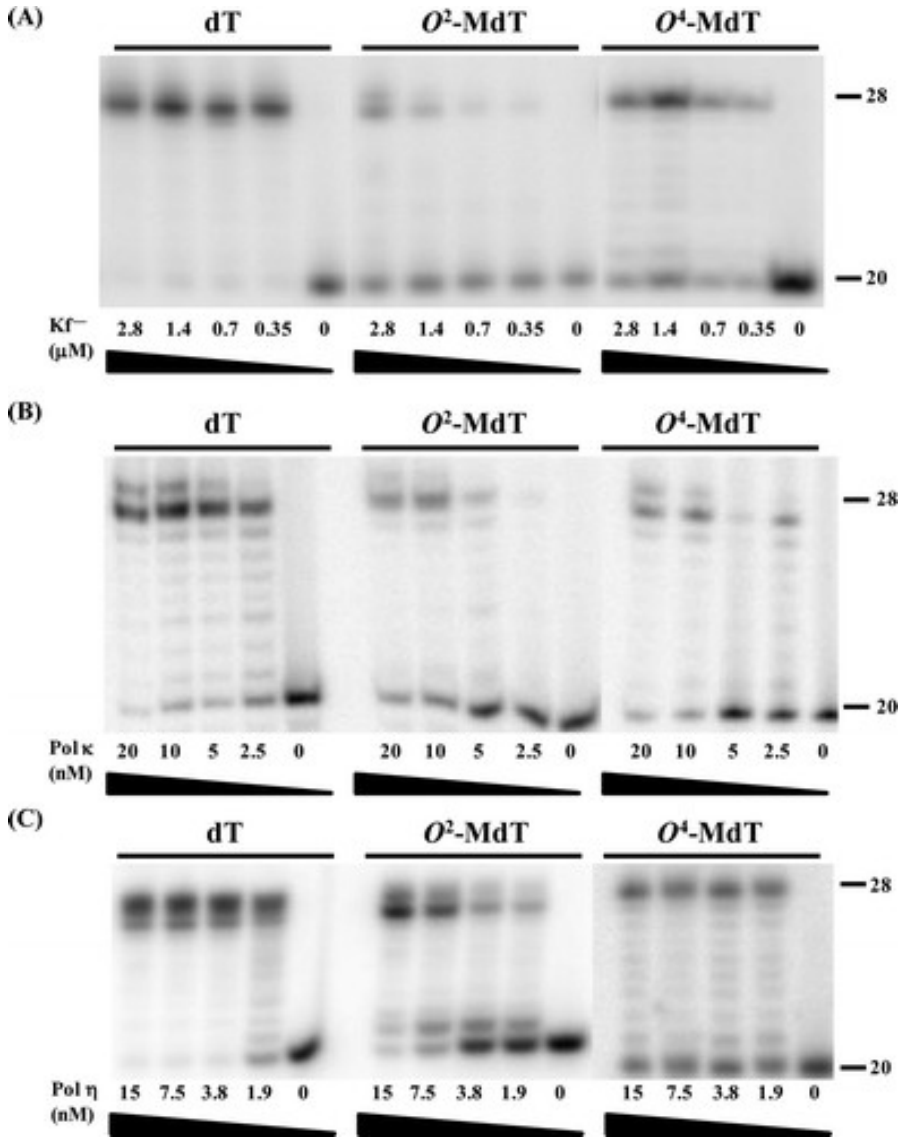
To correct for the varied ionization efficiencies of different ODNs, we constructed calibration curves using mixtures with varying concentrations of the standard synthetic ODNs identified in the reaction mixtures and a constant amount of the d(AATTCTCATGC) (11A, which represents the fully extended, unmutated product). Areas were determined for the peaks found in the selected-ion chromatogram (SIC) for monitoring the formation of three abundant fragment ions for each ODN. The peak areas of individual ODNs were then normalized to that of the 11A and plotted against the molar ratios of these ODNs over 11A to give calibration curves (Figure A.2). The corresponding normalized ratios for the replication samples were also determined, from which we measured the molar ratios for each extended product over 11A based on the calibration curves. The percentage of each product was then calculated based on the molar ratios of all products detected in the replication mixture.

Results

Primer Extension Assay Monitored by PAGE

To examine the effects of O^2-MdT and O^4-MdT on DNA replication, we performed *in vitro* replication studies using three DNA polymerases, the exonuclease-free Klenow fragment of *E. coli* DNA polymerase I, human DNA polymerase κ and *S. cerevisiae* DNA polymerase η. The results showed that, while full-length products were observed with the use of high concentrations of the enzymes, O^2-MdT and O^4-MdT are moderately blocking to all three polymerases (Figure 2.2).

Figure 2.2. Primer extension assays for O^2-MdT- and O^4-MdT-bearing substrates and the control undamaged substrate with Kf⁻ (A), human polymerase κ (B), and yeast polymerase η (C). The sequences for the templates are d(CCATGGCA*X*GAGAATTCTATGATCCTAG) ('*X*' represents dT, O^2-MdT or O^4-MdT), and a 5'-[^{32}P]-labeled d(GCTAGGATCATAGAATTCTC) was used as the primer.



Steady-state Kinetic Studies

Steady-state kinetic parameters for Kf⁻-, pol κ- and pol η-mediated nucleotide incorporation opposite O^2-MdT, O^4-MdT and corresponding unmodified dT in the above substrates were determined. Relative to the unmodified substrate, the efficiencies for Kf⁻ to incorporate the correct dAMP opposite O^2-MdT and O^4-MdT were reduced, with the k_{cat}/K_m values being 140, 5.7, and 2.0×10^{-2} μM^{-1} min^{-1} for substrates containing dT, O^2-MdT and O^4-MdT, respectively (Table 2.1, Figure A.3). Thus, methylation at the O^2- and O^4-positions of thymidine confers a reduction in efficiency for dAMP insertion by approximately 25 and 7000 folds, respectively. The magnitude of the reduction in efficiency for dAMP insertion opposite O^2-MdT relative to unmodified dT is less than what was previously reported,[33] which may be attributed to the different sequence contexts used for this and previous studies; sequence context is known to affect the efficiencies of nucleotide insertion.[28] Kf⁻ incorporates the other three nucleotides opposite O^2-MdT at much lower efficiencies than dAMP, i.e., 200 times lower for dCMP and dTMP, and 3 orders of magnitude lower for dGMP (Table 2.1). On the other hand, we found that Kf⁻ incorporates dAMP, dCMP and dTMP opposite O^4-MdT at similar efficiencies, though the polymerase displays a much stronger preference (by ~100-fold) for incorporation of dGMP over dAMP (Table 2.1). Together, Kf⁻ preferentially inserts the correct dAMP opposite O^2-MdT, but the incorrect dGMP opposite O^4-MdT.

Table 2.1. Efficiency and fidelity of *E. coli* Kf$^-$-mediated nucleotide incorporation opposite undamaged dT, O^2-MdT and O^4-MdT as determined by steady-state kinetic measurements.*

dNTP	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (μM^{-1} min^{-1})	f_{inc}
		Undamaged dT-containing Substrate		
dATP	2.6 ± 0.5	0.018 ± 0.001	140	1
dCTP	3.9 ± 0.3	160 ± 30	2.4×10^{-2}	1.7×10^{-4}
dGTP	5.5 ± 0.8	150 ± 20	3.7×10^{-2}	2.5×10^{-4}
dTTP	5.5 ± 0.6	490 ± 30	1.1×10^{-2}	7.8×10^{-5}
		O^2-MdT-containing Substrate		
dATP	2.5 ± 0.3	0.44 ± 0.07	5.7	1
dCTP	2.9 ± 0.4	130 ± 50	2.2×10^{-2}	3.9×10^{-3}
dGTP	3.5 ± 0.8	24000 ± 2000	1.5×10^{-4}	2.6×10^{-5}
dTTP	5.0 ± 0.6	2300 ± 200	2.2×10^{-3}	3.8×10^{-4}
		O^4-MdT-containing Substrate		
dATP	5.1 ± 0.4	260 ± 40	2.0×10^{-2}	1
dCTP	2.6 ± 0.2	230 ± 10	1.1×10^{-2}	0.58
dGTP	5.2 ± 0.3	1.9 ± 0.1	2.7	140
dTTP	3.3 ± 0.1	1100 ± 100	3.0×10^{-3}	0.15

*The K_m and k_{cat} were average values based on three independent measurements

Relative to the unmodified substrate, the efficiencies for human pol κ to incorporate the correct nucleotide, dAMP, opposite O^2-MdT and O^4-MdT were diminished by ~270 and 460 fold, respectively, with the k_{cat}/K_m values being 210, 0.77, and 0.46 μM^{-1} min^{-1} for dT, O^2-MdT, and O^4-MdT, respectively (Table 2.2 and Figure A.4). The efficiencies for human pol κ to incorporate three other nucleotides, i.e., dGMP, dCMP, dTMP opposite unmodified dT were much lower than that for dAMP insertion (Table 2.2). Incorporation efficiencies for the O^2-MdT- and O^4-MdT-containing templates displayed only slight decreases, i.e., by 2-5 fold, for the other three nucleotides (Table 2.2). On the other hand, the efficiencies for the insertion of dCMP and dTMP opposite O^4-MdT were 2 fold lower than that for dAMP incorporation. These results demonstrate that human pol κ-mediated nucleotide incorporation opposite O^2-MdT and O^4-MdT is error-prone.

Table 2.2. Efficiency and fidelity of human polymerase κ-mediated nucleotide incorporation opposite undamaged dT, O^2-MdT and O^4-MdT as determined by steady-state kinetic measurements.*

dNTP	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (μM^{-1} min^{-1})	f_{inc}
		Undamaged dT-containing Substrate		
dATP	5.9 ± 0.2	0.028 ± 0.002	210	1
dCTP	4.4 ± 0.1	0.66 ± 0.07	6.7	0.032
dGTP	4.3 ± 0. 4	6.0 ± 0.3	0.72	0.0034
dTTP	1.3 ± 0.2	63 ± 4	2.2×10^{-2}	9.5×10^{-5}
		O^2-MdT-containing Substrate		
dATP	2.8 ± 0.1	3.7 ± 0.2	0.76	1
dCTP	4.1 ± 0.1	10 ± 1	0.41	0.54
dGTP	3.0 ± 0.2	19 ± 1	0.16	0.21
dTTP	3.8 ± 0.6	21 ± 3	0.18	0.24
		O^4-MdT-containing Substrate		
dATP	1.8 ± 0.7	4.0 ± 0.1	0.45	1
dCTP	2.9 ± 0.1	15 ± 1	0.19	0.43
dGTP	4.0 ± 0.3	34 ± 2	0.12	0.26
dTTP	2.9 ± 0.7	14 ± 1	0.21	0.46

*The K_m and k_{cat} were average values based on three independent measurements

The efficiencies for yeast pol η to incorporate the correct nucleotide, dAMP, opposite dT, O^2-MdT, and O^4-MdT were 180, 0.34, and 7.6×10^{-3} μM^{-1} min^{-1}, respectively (Table 2.3 and Figure A.5). Interestingly, yeast pol η inserts dGMP (0.57 μM^{-1} min^{-1}) opposite O^4-MdT ~80 times more efficiently than dAMP (7.6×10^{-3} μM^{-1} min^{-1}, Table 2.3); the incorporation efficiencies for dCMP and dTMP were, however, 2-3 times lower than that for dAMP insertion. By contrast, we observed lower efficiencies for the incorporation of incorrect nucleotides opposite O^2-MdT; relative to dAMP insertion, the efficiency was approximately 10 times lower for dGMP and 20-30 times lower for dCMP and dTMP (Table 2.3). These results support that yeast pol η-mediated nucleotide incorporation opposite O^2-MdT in the template strand is accurate, whereas the template O^4-MdT primarily directs dGMP misincorporation.

Table 2.3. Efficiency and fidelity of yeast polymerase η-mediated nucleotide incorporation opposite undamaged dT, O^2-MdT and O^4-MdT as determined by steady-state kinetic measurements.*

dNTP	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (μM^{-1} min^{-1})	f_{inc}
		Undamaged dT-containing Substrate		
dATP	1.8 ± 0.2	0.01 ± 0.001	180	1
dCTP	1.2 ± 0.2	0.98 ± 0.03	1.2	6.7×10^{-3}
dGTP	2.1 ± 0.2	5.3 ± 0.4	0.39	2.2×10^{-3}
dTTP	3.0 ± 0.3	3.8 ± 0.3	0.80	4.4×10^{-3}
		O^2-MdT-containing Substrate		
dATP	2.6 ± 0.1	7.6 ± 0.2	0.34	1
dCTP	4.6 ± 0.1	230 ± 30	2.0×10^{-2}	5.8×10^{-2}
dGTP	8.5 ± 0.1	510 ± 10	1.7×10^{-2}	4.9×10^{-2}
dTTP	3.6 ± 0.1	310 ± 50	1.2×10^{-2}	3.4×10^{-2}
		O^4-MdT-containing Substrate		
dATP	2.2 ± 0.1	290 ± 70	7.6×10^{-3}	1
dCTP	2.5 ± 0.3	2100 ± 400	1.2×10^{-3}	1.6×10^{-1}
dGTP	0.85 ± 0.07	1.5 ± 0.03	0.57	75
dTTP	3.3 ± 0.4	840 ± 50	3.9×10^{-3}	5.2×10^{-1}

*The K_m and k_{cat} were average values based on three independent measurements

Primer Extension Studies with the use of LC-MS/MS

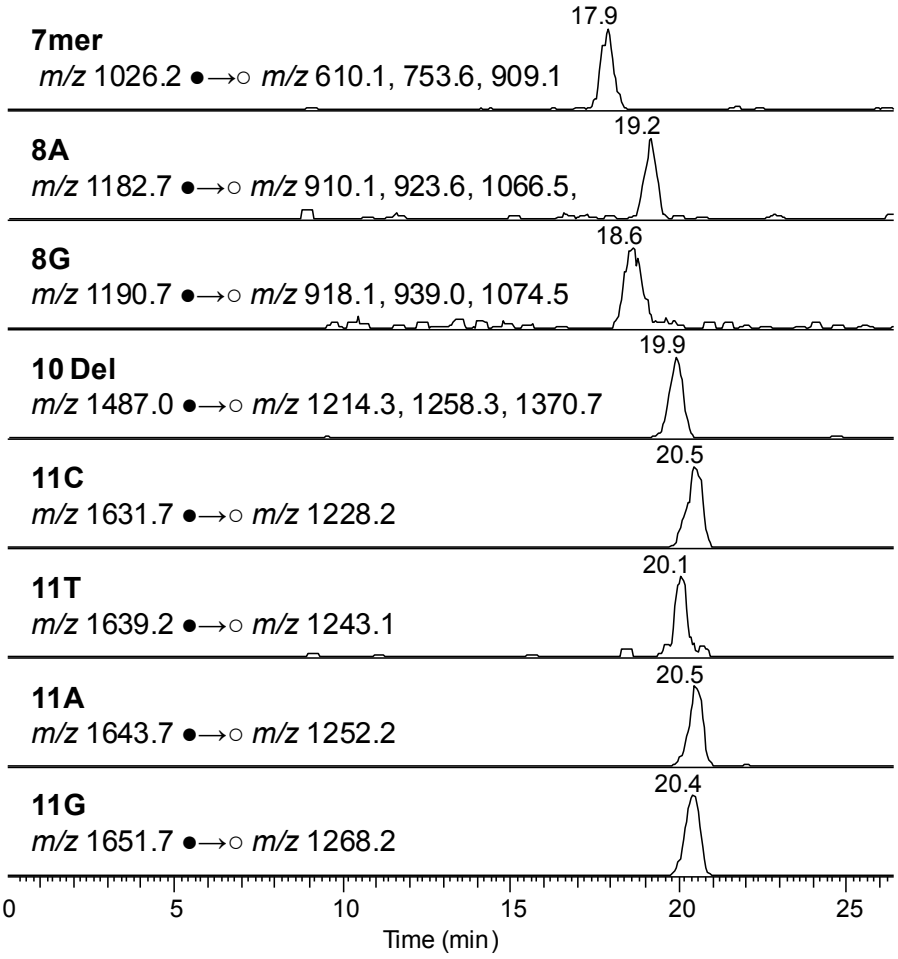
The above steady-state kinetic analysis provides useful information about how O^2- and O^4-MdT direct DNA polymerases to insert nucleotides opposite these lesions. However, such analysis may not reflect faithfully the nucleotide incorporation conditions in cells where DNA synthesis occurs in the presence of all four canonical nucleotides. Additionally, the primers carrying the correct or wrong nucleotide opposite the lesions may be extended at different efficiencies.[44] Thus, we employed LC-MS/MS to interrogate the extension products following previously described methods with some modifications.[45, 46] In this regard, primer extension reactions were conducted using the same primer/template complex as described above except that the primer was not radiolabeled. Instead of employing a uracil-containing primer, which can be subsequently cleaved using uracil DNA glycosylase followed with hot piperidine treatment,[45, 46] we digested the reaction mixtures with two restriction enzymes to give shorter extension products for LC-MS/MS analysis (Scheme 2.1).

Similar as what we reported previously,[47] the LC-MS/MS results showed a detection limit of ~10 fmol for the 7mer, 8A, 8G, 10 Del, 11T, 11C, and 11G at a signal-to-noise ratio greater than 20. Generally, when operated in the data-dependent scan mode, the LC-MS/MS method allows for the detection of species present at 1% or above when a 10-pmol reaction mixture was injected for LC-MS/MS analysis.

LC-MS/MS analysis of the extension products for both O^2-MdT- and O^4-MdT-containing substrates revealed the presence of the unextended or incompletely extended primer along with full-length replication products. Here we use the human pol κ-

catalyzed primer extension of the O^2-MdT-carrying substrate as an example to illustrate how we employ LC-MS/MS for the identification and quantification of replication products. The total-ion chromatogram (Figure A.6) reveals the 7mer unextended primer d(AATTCTC), the +1 products (8A and 8G), the 10mer deletion product, and the full-length extension products (including 11C, 11T, 11A and 11G) eluting at 17.9, 19.2, 19.9 and 20.6 min, respectively (Figure 2.3). The digested damage-containing template d(CCATGGCA*X*GAG), where '*X*' is O^2-MdT, and the 5' portion of the original primer d(GCTAGGATCATAG) also elute at 20.6 min (ESI-MS averaged from this retention time is shown in Figure A.6, and the sequences for the identified products are listed in Table 2.4). The identities of the aforementioned ODNs were established from ESI-MS and MS/MS analyses (MS/MS shown in Figures A.7 and A.8). Using the same LC-MS/MS analysis, we were able to identify the replication products arising from the other eight *in vitro* replication reactions (Table 2.4). It is worth noting that the LC-MS/MS results revealed that only ~5% of the O^2-MdT was degraded to the corresponding unmodified dT-containing substrate, supporting that the damage-containing substrate remains largely intact during the primer extension and restriction digestion conditions (Figures A.9 and A.10).

Figure 2.3. Selected-ion chromatograms obtained from the LC-MS and MS/MS analysis of the human pol κ-induced replication products that have been treated with two restriction enzymes, NcoI and EcoRI together with shrimp alkaline phosphatase. A 10-pmol replication mixture was injected for analysis.



5′- GC A**X**G AGA ATT -3′		
3′- C TCT TAA -5′	**7mer**	**6%**
3′- AC TCT TAA -5′	**8A**	**6%**
3′- GC TCT TAA -5′	**8G**	**8%**
3′- CG T C TCT TAA -5′	**10Del**	**3%**
3′- CG TCC TCT TAA -5′	**11C**	**5%**
3′- CG TTC TCT TAA -5′	**11T**	**1%**
3′- CG TAC TCT TAA -5′	**11A**	**26%**
3′- CG TGC TCT TAA -5′	**11G**	**46%**

O^2-MdT Replication From pol κ

Table 2.4. Summary of the percentages of replication products produced for dT-, O^2-MdT- and O^4-MdT-containing substrates as determined by LC-ESI-MS/MS experiments.

Name	Sequence	dT	O^2-MdT	O^4-MdT
	Klenow Fragment (exo⁻)			
7mer	d(AATTCTC)	2	2	7
8mer	d(AATTCTCA)		9	
8mer	d(AATTCTCG)		4	5
11A	d(AATTCTCATGC)	98	77	7
11G	d(AATTCTCGTGC)		7	80
	Yeast Polymerase η			
7mer	d(AATTCTC)	4	6	6
8mer	d(AATTCTCA)		2	
8mer	d(AATTCTCG)		6	2
10Del	d(AATTCTCTGC)		1	
11C	d(AATTCTCCTGC)		5	
11T	d(AATTCTCTTGC)		6	
11A	d(AATTCTCATGC)	96	45	9
11G	d(AATTCTCGTGC)		27	83
	Human Polymerase κ			
7mer	d(AATTCTC)	1	6	11
8mer	d(AATTCTCA)		6	
8mer	d(AATTCTCG)		8	18
10Del	d(AATTCTCTGC)		3	
11C	d(AATTCTCCTGC)		5	2
11T	d(AATTCTCTTGC)		1	3
11A	d(AATTCTCATGC)	99	26	8
11G	d(AATTCTCGTGC)		46	58

The template was d(CCATGGCA***X***GAGAATTCTATGATCCTAG), where ***'X'*** represents dT, O^2-MdT or O^4-MdT.

Extension products arising from replication by all three DNA polymerases opposite the undamaged template resulted in, as expected, a single extension product with the insertion of the correct nucleotide (i.e., dAMP) opposite the unmodified dT (Table 2.4). On the other hand, LC-MS/MS analysis of primer extension products for substrates containing O^2- and O^4-MdT revealed the presence of unextended primer, incompletely extended primer (+1 product), and full-length extension products.

O^4-MdT was blocking to the DNA polymerases; the total amount of unextended primer and +1 product (8G) accounted for 12%, 8%, and 29% of all products identified in mixtures arising from replication by Kf$^-$, yeast pol η, and human pol κ, respectively. Full-length DNA synthesis mediated by Kf$^-$, yeast pol η and human pol κ resulted in the incorrect dGMP being predominately incorporated opposite O^4-MdT, with 11G accounting for 80%, 83% and 58% of all the identified replication products, respectively (Table 2.4). Extension products with the correct dAMP opposite O^4-MdT represent a relatively small proportion of full-length products from reactions with Kf$^-$ (7%), yeast pol η (9%) and human pol κ (8%). Consistent with our steady-state kinetic results, pol κ also induced full-length extension products with incorporation of dCMP (2%) and dTMP (3%) opposite O^4-MdT.

O^2-MdT was blocking to DNA synthesis mediated by all three DNA polymerases studied, which is reflected by the presence of more unextended primer and +1 extension products (8A and 8G, Table 2.4) than replication opposite O^4-MdT; the total amounts of unextended primer and +1 extension products represent 15%, 15%, and 23% of all products arising from replication by Kf$^-$, yeast pol η, and human pol κ, respectively.

However, all three polymerases were capable of producing full-length extension products. Kf$^-$ predominately yielded full-length product with the correct dAMP (77%) being inserted opposite O^2-MdT, while misincorporation of dGMP (7%) occurs at a moderate frequency. Yeast pol η also preferentially extended the primer with the correct dAMP (45%) opposite O^2-MdT, though full-length products with dCMP (5%), dTMP (6%) and dGMP (27%) being incorporated opposite the lesion were also observed. Nevertheless, human pol κ predominately produced full-length products with the incorrect dGMP (46%) being inserted opposite O^2-MdT, followed by dAMP (26%), dCMP (5%) and dTMP (1%). Interestingly, replication mediated by yeast pol η and human pol κ also resulted in a -1 deletion product at frequencies of 1% and 3%, respectively, where the polymerase skipped the O^2-MdT site and continued replication past the damaged nucleoside (Table 2.4, Figure 2.3).

It is worth noting that, in order to obtain a relatively large amount of full-length replication products and enable their facile identification by LC-MS/MS, the above primer extension reactions were carried out with relatively high concentrations of dNTPs (1 mM each) and for a relatively long period of time (overnight), which may deviate from cellular DNA replication conditions. To assess whether the dNTP concentration and reaction time affect significantly the distribution of reaction products, we also conducted the corresponding experiments for the pol η-mediated replication of the O^4-MdT-bearing substrate where the individual dNTP concentrations were lowered to a more biologically relevant level (50 μM each) or the reaction time was shortened to 6 hrs. Similar to what were observed for the reaction conditions described above, the use of 50 μM dNTPs or a

6-hr incubation time mainly yield full-length products with the misincorporation of dGMP opposite the lesion at 73% and 62%, respectively (Table A.1). On the other hand, both reaction conditions resulted in higher proportions of unextended primer, which represent 17% and 35% (for 6-hr reaction) of all identified products for the reactions with lower dNTP concentration and shorter reaction time, respectively (Table A.1). Together, these results suggest that the primer extension conditions described above are suitable for revealing the miscoding potential of the DNA lesions under investigation.

Differences in Preferences for Nucleotide Insertion as Revealed by Steady-state Kinetic Assay and LC-MS/MS Analysis of Primer Extension Products

The above results obtained from the steady-state kinetic assay and LC-MS/MS analysis of primer extension reaction mixture revealed some notable differences for pol κ-mediated replication of O^2-MdT- and O^4-MdT-containing substrates, and for pol η-mediated replication of O^2-MdT-bearing substrate (Tables 2.2-2.4). For instance, steady-state kinetic assay showed -that human pol κ preferentially inserted dAMP over dGMP opposite O^2-MdT (by ~4.8 fold) and O^4-MdT (by ~3.8 fold, Table 2.2). LC-MS/MS data, however, showed that the full-length replication products carry mainly a dGMP opposite the two lesions; 46% 11G and 26% 11A were observed for the O^2-MdT-containing substrate, while 58% 11G and 8% 11A were found for the substrate harboring an O^4-MdT (Table 2.4). Choi et al. [48] made a similar finding for pol η-mediated replication of an O^6-methylguanine-housing substrate; steady-state kinetic experiment shows a slightly more preferential incorporation of dCMP over dTMP, but LC-MS/MS analysis of

extension products revealed that dTMP is inserted opposite the lesion 3 times more preferentially than dCMP.

We reason that several factors may contribute to these differences. As stated previously, the steady-state kinetic experiments were performed in the presence of a single nucleotide, whereas replication reactions for LC-MS/MS analyses were performed in the presence of all four natural nucleotides. The selectivity in nucleotide insertion under the latter reaction condition may not be faithfully recapitulated in steady-state kinetic assay where there is an absence of competition for the insertion of different nucleotides. Second, the primer with dAMP or dGMP being inserted opposite the lesion may be extended at different efficiencies. To assess the degree to which the second factor contributes to the differences in findings made from the two different assays, we conducted steady-state kinetic experiments for pol κ to insert the correct nucleotide, dTMP, opposite the adjoining 5' nucleoside of the lesion (dA). It turned out that the primer with dG being placed opposite O^2-MdT is extended 1.4 times more efficiently than the corresponding primer with a dA opposite the lesion, whereas the primer with dG being situated across O^4-MdT is extended at an efficiency that is 5.2 fold more than the corresponding primer with a dA (Table A.2 and Figure A.11). Thus, the difference in extension efficiency alters the distribution of full-length products, which, however, does not fully account for the differences observed from the LC-MS/MS and steady-state kinetic assays. On the basis of this observation, we may also conclude that the relative efficiencies for nucleotide incorporation opposite a lesion, as revealed by steady-state

kinetic assay, may not reflect faithfully the selectivity in nucleotide insertion opposite the lesion in cells where all four natural nucleotides are simultaneously present.

Discussion

Alkylating agents are known to react extensively with DNA, producing mutagenic and carcinogenic adducts in a variety of prokaryotic and eukaryotic organisms.[25] O^2-MdT and O^4-MdT can be produced in cellular DNA upon treatment with the carcinogenic *N*-methyl-*N*-nitrosourea.[6] Despite representing a small percentage of the observed products (~0.1% of total alkylation *in vivo*), O^2-MdT and O^4-MdT are not easily repaired *in vivo*[6, 10] and therefore might contribute significantly to the observed nucleobase substitutions and deletions induced from exposure to DNA methylating agents.

O^4-MdT was found to be highly mutagenic *in vitro* and *in vivo*, producing mainly T→C transitions,[2, 6, 10, 26-29, 49] and our results are in agreement with these previous observations. Both Kf⁻ and yeast pol η exhibited a strong preference for incorporating the incorrect dGMP opposite O^4-MdT. This is in line with our recent observation that O^4-carboxymethylthymidine (O^4-CMdT) directed yeast pol η to insert preferentially the incorrect dGMP.[44] On the other hand, steady-state kinetic assay results showed that human pol κ predominately incorporated the correct dAMP, followed by dCMP and dTMP, opposite O^4-MdT (Table 2.2). LC-MS/MS analysis of the primer extension results unveiled that the full-length products yielded by all three DNA polymerases carry mainly a dG opposite the O^4-MdT (Table 2.4). This finding suggests that the placement of a dG opposite the lesion renders efficient DNA synthesis beyond the lesion site. This is

particularly true for human pol κ; although, among the four natural nucleotides, dGMP is inserted opposite O^4-MdT at the lowest efficiency (Table 2.2), LC-MS/MS analysis revealed that incorporation of a dG opposite the lesion was the most favored (Table 2.4). The facile incorporation of dGMP opposite O^4-MdT may be attributed to the fact that thymine with a 4-alkyl group has been proposed to adopt a wobble conformation with guanine, thus allowing O^4-MdT to code as a cytosine.[50] Taken together, these results suggest that O^4-MdT mainly induces T→C transitions with all three DNA polymerases studied.

Jasti et al.[32] showed recently that O^2-MdT constitutes a strong block to DNA synthesis in *E. coli* cells even under SOS conditions, and the lesion is capable of inducing targeted base substitutions and, to a small degree, frameshift mutations. Our results revealed that incorporation of nucleotides opposite, and extension past, O^2-MdT is relatively inefficient by Kf^-, yeast pol η, or human pol κ. Additionally, base substitutions and, to a small degree, a -1 deletion were observed for DNA synthesis mediated by both yeast pol η and human pol κ opposite O^2-MdT (Tables 2.2-2.4). This finding is in stark contrast with previous observations that several minor-groove N^2-substituted dG derivatives can be efficiently and accurately bypassed by DinB family polymerases.[15-18] X-ray crystal structure of the catalytic core of human pol κ in complex with the primer/template and incoming nucleotide revealed the lack of steric hindrance in the minor groove at the primer-template junction.[51] This structure feature may permit pol κ to accommodate O^2-MdT into the active site of the polymerase. However, the unique hydrogen bonding property of O^2-MdT may not favor its base pairing with any of the

four canonical nucleotides, thereby preventing efficient incorporation of any nucleotide opposite the lesion. The ability of O^2-MdT in stalling DNA replication and in directing nucleotide misincorporation, in conjunction with the poor repair of O^2-alkylthymidine lesions,[10, 52, 53] suggests that this lesion may bear significant biological consequences.

Although O^2-MdT and O^4-MdT only constitute a small fraction of DNA lesions produced by methylating agents, these lesions, if left unrepaired, may lead to mutations in the genome. Further investigation is needed for addressing whether other O^2-alkylthymidine lesions are capable of inhibiting DNA replication and inducing mutations *in vitro*, and how these lesions compromise DNA replication in cells.

References

(1) Tannenbaum, S. R. (1983) *N*-nitroso compounds: A perspective on human exposure. *Lancet*, *1*, 629-632.

(2) Preston, B. D., Singer, B., and Loeb, L. A. (1986) Mutagenic potential of O^4-methylthymine *in vivo* determined by an enzymatic approach to site-specific mutagenesis. *Proc. Natl. Acad. Sci. U S A*, *83*, 8501-8505.

(3) Beranek, D. T. (1990) Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutat. Res.*, *231*, 11-30.

(4) Singer, B., and Essigmann, J. M. (1991) Site-specific mutagenesis: Retrospective and prospective. *Carcinogenesis*, *12*, 949-955.

(5) Vogel, E. W., and Nivard, M. J. (1994) International commission for protection against environmental mutagens and carcinogens. The subtlety of alkylating agents in reactions with biological macromolecules. *Mutat. Res.*, *305*, 13-32.

(6) Singer, B. (1986) *O*-alkyl pyrimidines in mutagenesis and carcinogenesis: Occurrence and significance. *Cancer Res.*, *46*, 4879-4885.

(7) Zak, P., Kleibl, K., and Laval, F. (1994) Repair of O^6-methylguanine and O^4-methylthymine by the human and rat O^6-methylguanine-DNA methyltransferases. *J. Biol. Chem.*, *269*, 730-733.

(8) Tominaga, Y., Tsuzuki, T., Shiraishi, A., Kawate, H., and Sekiguchi, M. (1997) Alkylation-induced apoptosis of embryonic stem cells in which the gene for DNA-repair, methyltransferase, had been disrupted by gene targeting. *Carcinogenesis*, *18*, 889-896.

(9) Altshuler, K. B., Hodes, C. S., and Essigmann, J. M. (1996) Intrachromosomal probes for mutagenesis by alkylated DNA bases replicated in mammalian cells: A comparison of the mutagenicities of O^4-methylthymine and O^6-methylguanine in cells with different DNA repair backgrounds. *Chem. Res. Toxicol.*, *9*, 980-987.

(10) Dosanjh, M. K., Singer, B., and Essigmann, J. M. (1991) Comparative mutagenesis of O^6-methylguanine and O^4-methylthymine in *Escherichia coli*. *Biochemistry*, *30*, 7027-7033.

(11) Den Engelse, L., Menkveld, G. J., De Brij, R. J., and Tates, A. D. (1986) Formation and stability of alkylated pyrimidines and purines (including imidazole ring-opened 7-alkylguanine) and alkylphosphotriesters in liver DNA of adult rats treated with ethylnitrosourea or dimethylnitrosamine. *Carcinogenesis*, *7*, 393-403.

(12) Friedberg, E. C., Wagner, R., and Radman, M. (2002) Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science*, *296*, 1627-1630.

(13) Prakash, S., Johnson, R. E., and Prakash, L. (2005) Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function. *Annu. Rev. Biochem.*, *74*, 317-353.

(14) Nohmi, T. (2006) Environmental stress and lesion-bypass DNA polymerases. *Annu. Rev. Microbiol.*, *60*, 231-253.

(15) Choi, J. Y., Angel, K. C., and Guengerich, F. P. (2006) Translesion synthesis across bulky N^2-alkyl guanine DNA adducts by human DNA polymerase κ. *J. Biol. Chem.*, *281*, 21062-21072.
(16) Jarosz, D. F., Godoy, V. G., Delaney, J. C., Essigmann, J. M., and Walker, G. C. (2006) A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature*, *439*, 225-228.
(17) Yuan, B., Cao, H., Jiang, Y., Hong, H., and Wang, Y. (2008) Efficient and accurate bypass of N^2-(1-carboxyethyl)-2'-deoxyguanosine by DinB DNA polymerase *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. U S A*, *105*, 8679-8684.
(18) Yuan, B., You, C., Andersen, N., Jiang, Y., Moriya, M., O'Connor, T. R., and Wang, Y. (2011) The roles of DNA polymerases κ and ι in the error-free bypass of N^2-carboxyalkyl-2'-deoxyguanosine lesions in mammalian cells. *J. Biol. Chem.*, *286*, 17503-17511.
(19) Johnson, R. E., Prakash, S., and Prakash, L. (1999) Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Pol η. *Science*, *283*, 1001-1004.
(20) Washington, M. T., Johnson, R. E., Prakash, S., and Prakash, L. (2000) Accuracy of thymine-thymine dimer bypass by Saccharomyces cerevisiae DNA polymerase η. *Proc. Natl. Acad. Sci. U S A*, *97*, 3094-3099.
(21) Yoon, J. H., Prakash, L., and Prakash, S. (2009) Highly error-free role of DNA polymerase η in the replicative bypass of UV-induced pyrimidine dimers in mouse and human cells. *Proc. Natl. Acad. Sci. U S A*, *106*, 18219-18224.
(22) Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature*, *399*, 700-704.
(23) Johnson, R. E., Prakash, S., and Prakash, L. (1999) Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Pol eta. *Science*, *283*, 1001-1004.
(24) Cleaver, J. E. (2005) Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat. Rev. Cancer*, *5*, 564-573.
(25) Shrivastav, N., Li, D., and Essigmann, J. M. (2010) Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis*, *31*, 59-70.
(26) Saffhill, R. (1985) *In vitro* miscoding of alkylthymines with DNA and RNA polymerases. *Chem. Biol. Interact.*, *53*, 121-130.
(27) Menichini, P., Mroczkowska, M. M., and Singer, B. (1994) Enzyme-dependent pausing during *in vitro* replication of O^4-methylthymine in a defined oligonucleotide sequence. *Mutat. Res.*, *307*, 53-59.
(28) Rao, S., Chenna, A., Slupska, M., and Singer, B. (1996) Replication of O^4-methylthymine-containing oligonucleotides: Effect of 3' and 5' flanking bases on formation and extension of O^4-methylthymine-guanine basepairs. *Mutat. Res.*, *356*, 179-185.

(29) Encell, L. P., and Loeb, L. A. (2000) Enhanced *in vivo* repair of O^4-methylthymine by a mutant human DNA alkyltransferase. *Carcinogenesis*, *21*, 1397-1402.
(30) Grevatt, P. C., Solomon, J. J., and Bhanot, O. S. (1992) *In vitro* mispairing specificity of O^2-ethylthymidine. *Biochemistry*, *31*, 4181-4188.
(31) Gowda, A. S., Krishnegowda, G., Suo, Z., Amin, S., and Spratt, T. E. (2012) Low fidelity bypass of O^2-(3-pyridyl)-4-oxobutylthymine, the most persistent bulky adduct produced by the tobacco specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone by model DNA polymerases. *Chem. Res. Toxicol.*, *25*, 1195-1202.
(32) Jasti, V. P., Spratt, T. E., and Basu, A. K. (2011) Tobacco-specific nitrosamine-derived O^2-alkylthymidines are potent mutagenic lesions in SOS-induced *Escherichia coli*. *Chem. Res. Toxicol.*, *24*, 1833-1835.
(33) Prakash G.A.S., K., G., Suo, Z., Amin, S. and Spratt, T. (2012) Low fidelity bypass of O^2-(3-pyridyl)-4-oxobutylthymine, the most persistent bulky adduct produced by the tobacco specific nitrosamine 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone by model DNA polymerases. *Biochemistry*, *25*, 1195-1202.
(34) Grevatt, P. C., Solomon, J.J. and Bhanot, O.S. (1992) *In vitro* mispairing specificity of O^2-ethylthymidine. *Biochemistry*, *31*, 4181-4188.
(35) Haracska, L., Prakash, S., and Prakash, L. (2000) Replication past O^6-methylguanine by yeast and human DNA polymerase η. *Mol. Cell. Biol.*, *20*, 8001-8007.
(36) Cannistraro, V. J., and Taylor, J. S. (2004) DNA-thumb interactions and processivity of T7 DNA polymerase in comparison to yeast polymerase η. *J. Biol. Chem.*, *279*, 18288-18295.
(37) Kuznetsov, N. A., Koval, V. V., Zharkov, D. O., Nevinsky, G. A., Douglas, K. T., and Fedorova, O. S. (2005) Kinetics of substrate recognition and cleavage by human 8-oxoguanine-DNA glycosylase. *Nucleic Acids Res.*, *33*, 3919-3931.
(38) Xu, Y. Z., and Swann, P. F. (1994) Oligodeoxynucleotides containing O^2-alkylthymine: Synthesis and characterization. *Tetrahedron Lett.*, *35*, 303-306.
(39) Gu, C., and Wang, Y. (2004) LC-MS/MS identification and yeast polymerase η bypass of a novel γ-irradiation-induced intrastrand cross-link lesion G[8-5]C. *Biochemistry*, *43*, 6745-6750.
(40) Lao, Y. B., Yu, N. X., Kassie, F., Villalta, P. W., and Hecht, S. S. (2007) Analysis of pyridyloxobutyl DNA adducts in F344 rats chronically treated with *R*- and *S*-*N*'-nitrosonornicotine. *Chem. Res. Toxicol.*, *20*, 246-256.
(41) Den Engelse, L., De Graaf, A., De Brij, R. J., and Menkveld, G. J. (1987) O^2- and O^4-ethylthymine and the ethylphosphotriester dTp(Et)dT are highly persistent DNA modifications in slowly dividing tissues of the ethylnitrosourea-treated rat. *Carcinogenesis*, *8*, 751-757.
(42) Creighton, S., Bloom, L. B., and Goodman, M. F. (1995) Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol.*, *262*, 232-256.

(43) Boosalis, M. S., Petruska, J., and Goodman, M. F. (1987) DNA-polymerase insertion fidelity: Gel assay for site-specific kinetics. *J. Biol. Chem.*, *262*, 14689-14696.

(44) Swanson, A. L., Wang, J., and Wang, Y. (2011) *In vitro* replication studies of carboxymethylated DNA lesions with *Saccharomyces cerevisiae* polymerase η. *Biochemistry*, *50*, 7666-7673.

(45) Gu, C., and Wang, Y. (2007) *In vitro* replication and thermodynamic studies of methylation and oxidation modifications of 6-thioguanine. *Nucleic Acids Res.*, *35*, 3693-3704.

(46) Zang, H., Goodenough, A. K., Choi, J. Y., Irimia, A., Loukachevitch, L. V., Kozekov, I. D., Angel, K. C., Rizzo, C. J., Egli, M., and Guengerich, F. P. (2005) DNA adduct bypass polymerization by *Sulfolobus solfataricus* DNA polymerase Dpo4: Analysis and crystal structures of multiple base pair substitution and frameshift products with the adduct 1,N^2-ethenoguanine. *J. Biol. Chem.*, *280*, 29750-29764.

(47) Hong, H., Cao, H., and Wang, Y. (2007) Formation and genotoxicity of a guanine-cytosine intrastrand cross-link lesion *in vivo*. *Nucleic Acids Res.*, *35*, 7118-7127.

(48) Choi, J. Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L. A., and Guengerich, F. P. (2006) Translesion synthesis across O^6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J. Biol. Chem.*, *281*, 38244-38256.

(49) Dosanjh, M. K., Essigmann, J. M., Goodman, M. F., and Singer, B. (1990) Comparative efficiency of forming m^4T-G versus m^4T-A base pairs at a unique site by use of *Escherichia coli* DNA polymerase I (Klenow fragment) and *Drosophila melanogaster* polymerase α-primase complex. *Biochemistry*, *29*, 4698-4703.

(50) Swann, P. F. (1990) Why do O^6-alkylguanine and O^4-alkylthymine miscode? The relationship between the structure of DNA containing O^6-alkylguanine and O^4-alkylthymine and the mutagenic properties of these bases. *Mutat. Res.*, *233*, 81-94.

(51) Lone, S., Townson, S. A., Uljon, S. N., Johnson, R. E., Brahma, A., Nair, D. T., Prakash, S., Prakash, L., and Aggarwal, A. K. (2007) Human DNA polymerase κ encircles DNA: implications for mismatch extension and lesion bypass. *Mol. Cell*, *25*, 601-614.

(52) McCarthy, T. V., Karran, P., and Lindahl, T. (1984) Inducible repair of *O*-alkylated DNA pyrimidines in *Escherichia coli*. *EMBO J.*, *3*, 545-550.

(53) Brent, T. P., Dolan, M. E., Fraenkelconrat, H., Hall, J., Karran, P., Laval, F., Margison, G. P., Montesano, R., Pegg, A. E., Potter, P. M., Singer, B., Swenberg, J. A., and Yarosh, D. B. (1988) Repair of *O*-alkylpyrimidines in mammalian cells: A present consensus. *Proc. Natl. Acad. Sci. U S A*, *85*, 1759-1762.

Chapter 3: Replication across Regioisomeric Ethylated Thymidine Lesions by Purified DNA Polymerases

Introduction

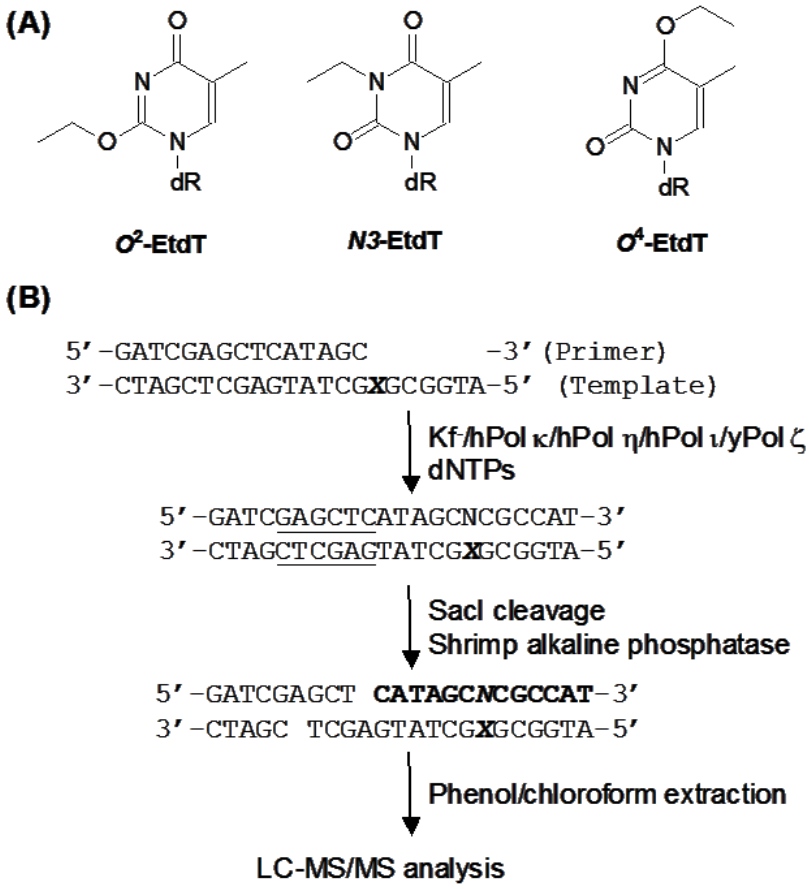
Exposure to tobacco smoke is thought to contribute to cardiovascular and pulmonary diseases and approximately 30% of all cancer deaths in developed countries.[1] Causal links exist between smoking cigarettes and human cancers of the oral cavity, pharynx, larynx, esophagus, pancreas, urinary bladder and renal pelvis, where over 80% of all lung cancers could be attributed to tobacco smoke exposure.[2] Cigarette smoke contains over 6000 compounds, many of which are genotoxic.[3] Carcinogens in tobacco smoke such as aromatic amines, polycyclic aromatic hydrocarbons and tobacco-specific nitrosamines require metabolic activation by cytochrome P450 family of enzymes for the formation of reactive, electrophilic species that can react with DNA to yield covalent adducts.[4]

Treatment with ethylating agents has been shown to induce ethylation products on backbone phosphate and all four nucleobases of DNA.[5] Exposure of DNA to cigarette smoke *in vitro* resulted in a dose-dependent increase of *N*7-ethylguanine (*N*7-EtG), though the identity and source of the direct-acting ethylating agent(s) in cigarette smoke has not been firmly established.[6] It has been proposed that ethylation of biomolecules may serve as biomarkers for monitoring exposure to ethylating agents in cigarette smoke and for cancer risk assessment. Along these lines, significantly elevated levels of N-terminal *N*-ethylvaline were found in hemoglobin from smokers relative to non-smokers.[7]

In subsequent studies, the levels of *N*3-ethyladenine (*N*3-EtA)[8, 9] and *N*7-EtG in urine,[10] O^4-ethylthymidine (O^4-EtdT) in lung tissue,[11] and *N*7-EtG in human leukocyte DNA[12] were significantly higher in smokers than non-smokers.

Recently, Chen et al.[13] quantified O^2-, *N*3- and O^4-ethylthymidine (O^2-EtdT, *N*3-EtdT, and O^4-EtdT) in human leukocyte DNA, where the levels of the three ethylated thymidine lesions were signicantly higher in smokers than non-smokers. Compared to other ethylated DNA lesions such as O^6-ethyl-2'-deoxyguanosine (O^6-EtdG, with a half-life of ~14 hrs in rat tissue), both O^2-EtdT and O^4-EtdT are poorly repaired and thus are highly persistent in mammalian tissues (half-life of ~11-20 days in rat tissue).[14-16] In addition, several studies have shown that DNA synthesis opposite O^2-, *N*3- and O^4-alkylthymidine is both blocking and mutagenic. [17-24] Here we investigated how site-specifically incorporated, regioisomeric O^2-, *N*3- and O^4-EtdT (Figure 3.1) affect DNA replication mediated by purified human DNA polymerases (hPol) η, κ, and ι, yeast DNA polymerase ζ (yPol ζ), and the exonuclease-free Klenow fragment of *E. coli* DNA polymerase I (Kf^-).

Figure 3.1. Structures of the regioisomeric ethylated thymidines. "dR" refers to 2-deoxyribose (A). Experimental procedures for monitoring primer extension products using SacI cleavage and LC-MS/MS (B). '*X*' represents dT, O^2-EtdT, *N*3-EtdT or O^4-EtdT, and 'N' designates the nucleotide incorporated opposite X during *in-vitro* replication.



Materials and Methods Materials

All enzymes and chemicals unless otherwise specified were purchased from New England Biolabs (Ipswich, MA) or Sigma (St. Louis, MO). Unmodified oligodeoxyribonucleotides (ODNs) used in this study were from Integrated DNA Technologies (Coralville, IA). [γ-^{32}P]ATP was obtained from Perkin Elmer (Piscataway, NJ). 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was purchased from TCI America (Portland, OR). Phosphoramidites for unmodified nucleosides and other reagents for solid-phase DNA synthesis were obtained from Glen Research (Sterling, VA). hPol κ and yPol ζ were purchased from Enzymax (Lexington, KY), hPol ι [25] and hPol η [26] were kindly provided by Drs. Roger Woodgate and Wei Yang (NIH, Bethesda, MD), respectively.

Chemical synthesis of the phosphoramidite building block of *N*3-EtdT

The procedures for the synthesis of the phosphoramidite building block of *N*3-EtdT were adapted from previously described procedures for the preparation of the *N*3-carboxyethylthymidine phosphoramidite [27].

Synthesis of N3-EtdT

In a round bottom flask, thymidine (242 mg, 1.00 mmol) and potassium carbonate (166 mg, 1.20 mmol) were dissolved in anhydrous methanol (50 mL). To the solution was subsequently added bromoethane (129 mg, 1.20 mmol), and the resulting mixture was refluxed for 6 h. The solvent was removed under reduced pressure and the residue purified by column chromatography using 15% methanol in ethyl acetate as mobile phase to give **1** as a white foam (210 mg, 78% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.40 (s,

1H), 6.22 (t, J = 6.8 Hz, 1H), 4.56 (dt, J = 5.8, 3.9 Hz, 1H), 4.03-3.93 (m, 3H), 3.86 (ddd, J = 24.6, 11.8, 3.1 Hz, 2H), 2.98 (s, 2H), 2.43-2.28 (m, 2H), 1.91 (s, 3H), 1.19 (t, J = 7.0 Hz, 3H). HRMS (ESI-TOF) calcd for $C_{12}H_{19}N_2O_5$ $[M+H]^+$ 271.1294, found 271.1314.

Synthesis of 5'-O-(4,4'-dimethoxytrityl)-N3-ethylthymidine

Compound **1** was dissolved in anhydrous pyridine (10 mL) and the solution was cooled in an ice bath, to which solution were added 4-dimethylaminopyridine (DMAP, 0.5% mol) and dimethoxytrityl chloride (DMTr-Cl, 1.2 eq.). The resulting solution was stirred at room temperature for 10 hrs. The reaction was then quenched with methanol (0.5 mL) and the solvent removed under reduced pressure. The residue was purified by silica gel column chromatography with ethyl acetate as mobile phase to yield **2** as a white foam (yield 76%): ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.55 (s, 1H), 7.43-7.21 (m, 9H), 6.83 (d, J = 8.8 Hz, 4H), 6.45 (t, J = 6.7 Hz, 1H), 4.56 (s, 1H), 4.07-3.94 (m, 3H), 3.79 (s, 6H), 3.51-3.35 (m, 3H), 2.52-2.25 (m, 3H), 1.51 (s, 3H), 1.21 (t, J = 6.9 Hz, 3H). HRMS (ESI-TOF) calcd for $C_{33}H_{36}N_2O_7Na$ $[M+Na]^+$ 595.2420, found 595.2401.

5'-O-(4,4'-dimethoxytrityl)-N3-ethylthymidine-3'-O-[(2-cyanoethyl)-N,N-diisopropyl-phosphoramidite]

To a round bottom flask, which was suspended in an ice bath and contained a solution of compound **2** in dry dichloromethane (3.0 mL), was added *N*,*N*-diisopropylethylamine (DIEA, 2.2 eq.) followed by dropwise addition of 2-cyanoethyl-*N*,*N*-diisopropyl chlorophosphoramidite (1.2 eq.). The mixture was stirred at room temperature for 1 hr under argon atmosphere. The reaction was quenched by cooling the

mixture in an ice bath followed by slow addition of methanol (0.40 mL). The solution was quickly diluted with ethyl acetate (8.0 mL). The organic layer was washed with saturated $NaHCO_3$ (4.0 mL) and brine (4.0 mL), and dried over anhydrous Na_2SO_4. The solvent was evaporated under reduced pressure to yield **3** in a foam that was used directly for ODN synthesis. **3**: ^{31}P NMR ($CDCl_3$): δ 150.01, 149.47.

NMR

^{1}H NMR and ^{31}P NMR spectra were acquired on a Varian Inova 300 NMR spectrometer (Varian Inc., Palo Alto, CA). Resonance assignments for *N*3-EtdT were made based on two-dimensional ^{1}H-^{13}C heteronuclear multi-bond correlation (HMBC) experiment. The HMBC spectrum was acquired a Varian Unity spectrometer operating at 500 MHz at 25°C using sweep widths of 3723.0 Hz and 30 177.3 Hz for ^{1}H and ^{13}C, respectively. The first delay was set to match a 140 Hz coupling constant, and the second delay was set to match a long-range coupling constant of 8 Hz.

ODN synthesis

The 12-mer lesion-containing ODNs 5'-ATGGCGXGCTAT-3' ('X' represents O^2-, *N*3-, or O^4-EtdT) were synthesized on a Beckman Oligo 1000S DNA synthesizer (Fullerton, CA) at 1-μmol scale. The phosphoramidite building blocks for O^2-EtdT and O^4-EtdT were synthesized following previously published procedures.[28, 29] The synthesized phosphoramidite building blocks of O^2-EtdT, *N*3-EtdT and O^4-EtdT were dissolved in anhydrous acetonitrile at a concentration of 0.067 M. Conventional phosphoramidite building blocks of unmodified nucleosides were employed, and a

standard ODN assembly protocol was used without any modification. The ODNs containing O^2-EtdT and O^4-EtdT were cleaved from the controlled pore glass (CPG) support with 10% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in ethanol in the dark at 23°C for five days, while *N*3-EtdT-containing ODN was deprotected with concentrated ammonium hydroxide at 55°C overnight. The ODNs were purified by HPLC and sequences verified by ESI-MS and MS/MS analyses (Figures B.1-B.3).

HPLC

HPLC separation of synthetic ODNs was performed on an Agilent 1100 HPLC system with an Aeris XB-C18 column (4.6×150 mm, 3.6 μm in particle size and 200 Å in pore size; Phenomenex Inc., Torrance, CA, USA). For the purification of ODNs, a triethylammonium acetate buffer (50 mM, pH 6.8, Solution A) and a mixture of solution A and acetonitrile (70/30, v/v, Solution B) were employed as mobile phases. The flow rate was 0.8 mL/min and the gradient included 0-25% B in 5 min, 25-50% B in 40 min, 50-95% B in 10 min and at 95% B for 5 min.

Preparation of the lesion-carrying 22-mer ODNs

The 22mer substrates, d(ATGGCG*X*GCTATGAGCTCGATC) ('*X*' = dT, O^2-EtdT, *N*3-EtdT or O^4-EtdT), were obtained by ligating the above-described 12mer lesion-containing or the corresponding dT-bearing ODNs with a 5'-phosphorylated d(GAGCTCGATC) in the presence of a template ODN following previously published procedures.[30] The desired lesion-containing 22mer ODNs were purified by PAGE and

desalted by ethanol precipitation. The purity of the products was further confirmed by PAGE analysis.

Primer extension assays monitored by gel electrophoresis

In-vitro replication experiments were performed following previously described procedures.[24, 31] Briefly, the 22mer template, d(ATGGCG*X*GCTATGAGCTCGATC) ('*X*' = dT, O^2-EtdT, *N*3-EtdT or O^4-EtdT, 20 nM), was annealed with a 5'-^{32}P-labeled 15mer primer, d(GATCGAGCTCATAGC) (10 nM). For Kf⁻, the reactions were carried out in a buffer containing 50 mM Tris-HCl (pH 7.5), 20 mM $MgCl_2$, 2 mM EDTA, 1.6 mM β-mercaptoethanol, and 5 μg/mL BSA. For hPol η, hPol κ and hPol ι, the reactions were conducted in a buffer containing 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, and 7.5 mM DTT. For primer extension assays, all four dNTPs and varying concentrations of a DNA polymerase, as indicated in the figures, were subsequently added to the duplex mixture and incubated for 60 min. The reactions were terminated by adding a 2-volume excess of formamide gel-loading buffer. The products were resolved on 20% (29:1) cross-linked polyacrylamide gels containing 8 M urea. Gel band intensities for the substrates and products were visualized by using a Typhoon 9410 variable-mode imager and data processed using ImageQuant 5.2 (GE Healthcare Life Sciences, Piscataway, NJ).

TLS cooperativity experiments were performed following previously described methods.[32] Briefly, the above-described primer/template complexes were incubated with hPol ι and all four dNTPs (100 μM) at 37°C for 30 min. hPol κ was subsequently added to the reaction mixture and incubated for an additional 30 min. These experiments were

conducted in a buffer containing 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, and 7.5 mM DTT. The products were resolved on 20% (29:1) cross-linked polyacrylamide gels containing 8 M urea and gel bands visualized as described above (Figure B.4).

Primer extension assays monitored by LC-MS/MS

The primer and lesion-containing templates (0.30 μM each) were annealed and incubated in a 50-μL solution containing hPol κ (0.10 μM), hPol η (0.10 μM), hPol ι (0.10 μM), yPol ζ (0.10 μM) or Kf⁻ (55 nM) at 37°C for 6 hr. The reactions were conducted in the presence of all four dNTPs (0.1 μM each) in the same buffers as described above, and terminated by removing the enzymes via chloroform extraction. The aqueous layer was dried using a Speed-Vac, and the dried residues redissolved in a buffer containing 50 mM NaCl, 10 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, and 1 mM DTT. SacI (40 U) and shrimp alkaline phosphatase (20 U) were subsequently added to the mixture and incubated at 37°C for 4 hr. The proteins in the mixture were again removed by chloroform extraction, and the aqueous layer dried. The dried residues were reconstituted in 30-μL H_2O, and a 10-μL aliquot was injected for LC-MS/MS analysis with the use of an Agilent Zorbax SB-C18 column (0.5 × 250 mm, 5 μm in particle size). The gradient for the HPLC elution was 5-min of 5-20% methanol followed by 35-min of 20-50% methanol in 400 mM HFIP (pH adjusted to 7.0). The temperature for the ion-transport tube was maintained at 275°C to minimize the formation of HFIP adducts of the ODNs.

To identify the replication products, samples were first analyzed in the data-dependent scan mode, where the most abundant ion found in MS was chosen for

fragmentation in MS/MS. The fragment ions observed in MS/MS were manually assigned and the sequences of the ODNs determined (Representative MS/MS for the identification of products from hPol η-mediated reaction of the O^2-EtdT-containing substrate is shown in Figure B.5). Following identification of the ODN products, the mass spectrometer was set up for monitoring specifically the fragmentation of the precursor ions for the extended portions of the primer strand.

To correct for the varied ionization efficiencies of different ODNs, we constructed calibration curves using mixtures with varying concentrations of the standard synthetic ODNs identified in the reaction mixtures and a constant amount of d(CATCGAGCT) (9mer 5'P, which was the 5' portion of the primer liberated from the SacI cleavage. Figures 3.1B and B.6). Areas were determined for the peaks found in the selected-ion chromatograms (SICs) by monitoring the formation of 1-3 unique and abundant fragment ions for each ODN. The peak areas of individual ODNs were then normalized to that of the 9mer 5'P and plotted against the molar ratios of these ODNs over 9mer 5'P to give the calibration curves (Figure B.6). The corresponding normalized ratios for the replication samples were also determined, from which we measured the molar ratios for each extended product over 9mer 5'P based on the calibration curves. The percentage of each product was then calculated from the molar ratios of all products detected in the replication mixture.

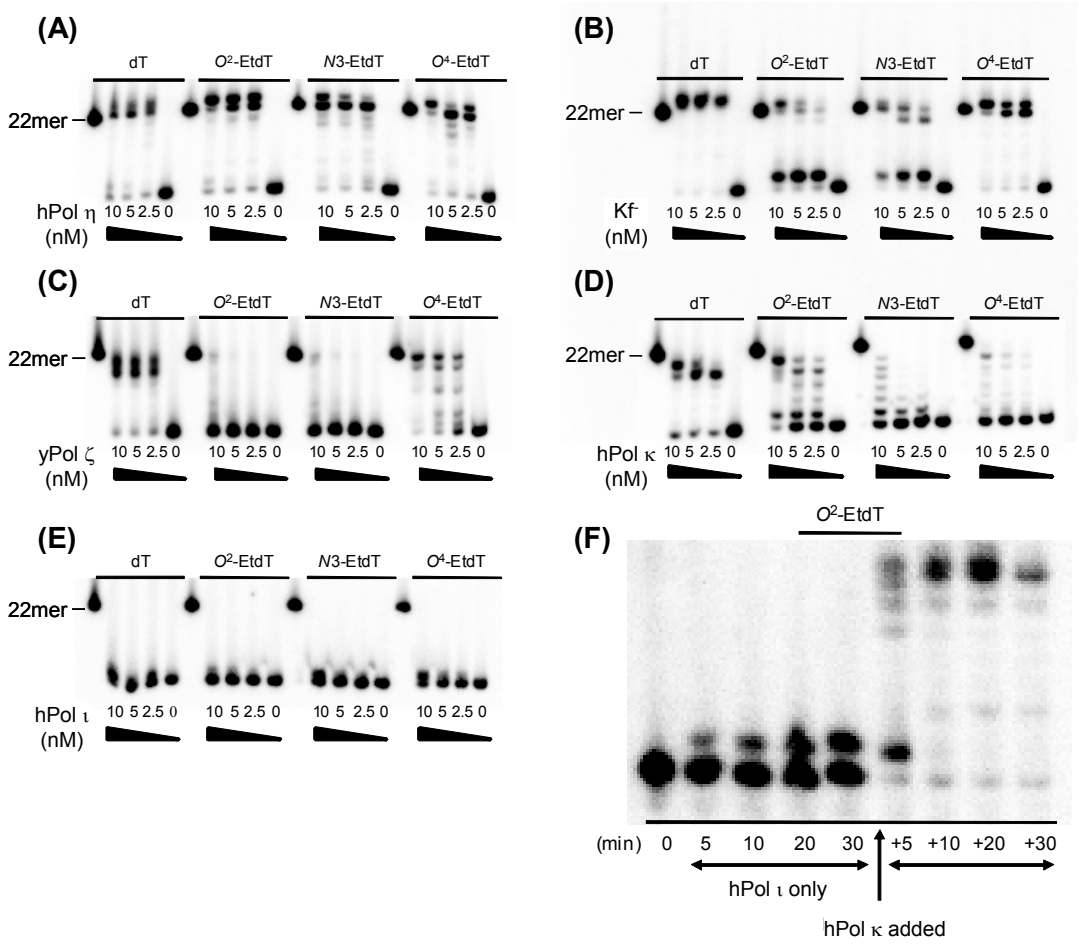
Results

The major objective of the present study was to assess systematically how the three site-specifically incorporated, regioisomeric O^2-, $N3$- and O^4-EtdT are recognized by DNA polymerases, including hPol η, hPol κ, hPol ι, yPol ζ, and Kf⁻. Kf⁻ was selected because it has been widely used as a model DNA polymerase for examining lesion bypass, and hPol η, hPol κ, hPol ι and yPol ζ were chosen because of their known abilities in bypassing a variety of DNA lesions. For example, hPol η and yPol ζ are involved in accurately bypassing *cis-syn* cyclobutane thymine dimers [33, 34], and hPol κ and hPol ι are both able to bypass the minor-groove N^2-substituted dG lesions.[35, 36]

Primer extension assay monitored by PAGE

Our primer extension assay results revealed that the three regioisomeric ethylated thymidine lesions are recognized and bypassed differently by these DNA polymerases (Figure 3.2). While hPol η was able to bypass readily all three ethylated dT lesions and generate full-length products, O^2-EtdT and $N3$-EtdT strongly blocked primer extension mediated by Kf⁻ and yPol ζ (Figure 3.2A-C). Kf⁻, for the most part, stalls after inserting one nucleotide opposite the two lesions, though full-length replication products were also observed for these two lesions. yPol ζ, on the other hand, was unable to insert any nucleotide opposite the two lesions (Figure 3.2B and C). hPol κ was strongly blocked by both $N3$-EtdT and O^4-EtdT; nevertheless, the polymerase was only moderately blocked by O^2-EtdT (Figure 3.2D). In this context, it is of note that blunt-end elongation products were also observed for the reactions catalyzed by Kf⁻ and hPol η (i.e., the 23-mer and

Figure 3.2. Primer extension assays for O^2-EtdT-, *N*3-EtdT- and O^4-EtdT-bearing substrates and the control undamaged substrate with hPol η (A), Kf$^-$ (B), yPol ζ (C), hPol κ (D), and hPol ι (E). (F) Cooperativity of human DNA polymerase ι and κ in the bypass of the O^2-EtdT lesion. Primer-template complex containing O^2-EtdT (50 nM) was incubated with hPol ι (10 nM) for 0-30 min. After 30 min, 5 nM of hPol κ was added to the reaction mixture to determine whether the primer extension started by hPol ι and stalled by the lesion can be completed by hPol κ. The sequences for the templates are d(ATGGCG*X*GCTATGAGCTCGATC) ('*X*' represents dT, O^2-EtdT, *N*3-EtdT or O^4-EtdT), and a 5'-[^{32}P]-labeled d(GATCGAGCTCATAGC) was used as the primer (see Figure 2.1B). "22mer" refers to a standard ODN for the fully extended primer strand.



24-mer products, Figure 3.2A and B), as observed previously.[37, 38] Similar as what were found previously,[38, 39] hPol ι was strongly blocked for the template containing an unmodified dT (Figure 3.2E); not surprisingly, the polymerase stalls after inserting one nucleotide opposite the three EtdT lesions (Figure 3.2E).

Additional experiments were carried out to determine if the incomplete extension products produced by hPol ι could be extended by hPol κ. It turns out that hPol κ was able to further extend the primer to yield full-length (20-mer) products for the templates containing O^2-EtdT, *N*3-EtdT, O^4-EtdT (Figures 3.2F and B.4), even though neither polymerase alone was able to extend the primer and generate full-length products for the three EtdT lesions. These results suggest the possible cooperation between hPol ι, which inserts one nucleotide opposite the ethylated lesions, and hPol κ, which extends the primer past the lesions.

Replication products monitored by LC-MS/MS

We next employed LC-MS/MS to identify and quantify the products emanating from the *in vitro* replication reactions following previously published methods with some modifications.[40, 41] Instead of using a uracil-containing primer, which could be subsequently treated with uracil DNA glycosylase and hot piperidine to give shorter products,[32, 38, 40] we employed a primer/template sequence which, after *in-vitro* replication, could be cleaved with a restriction enzyme (i.e., SacI). The SacI cleavage yields shorter ODN products that are amenable to LC-MS/MS identification and quantification (Figures 3.1B). For the quantification, we utilized the 5' portion of the

cleaved primer (9-mer 5'P) as a reference to quantify the relative amounts of replication products (See Experimental Procedures).

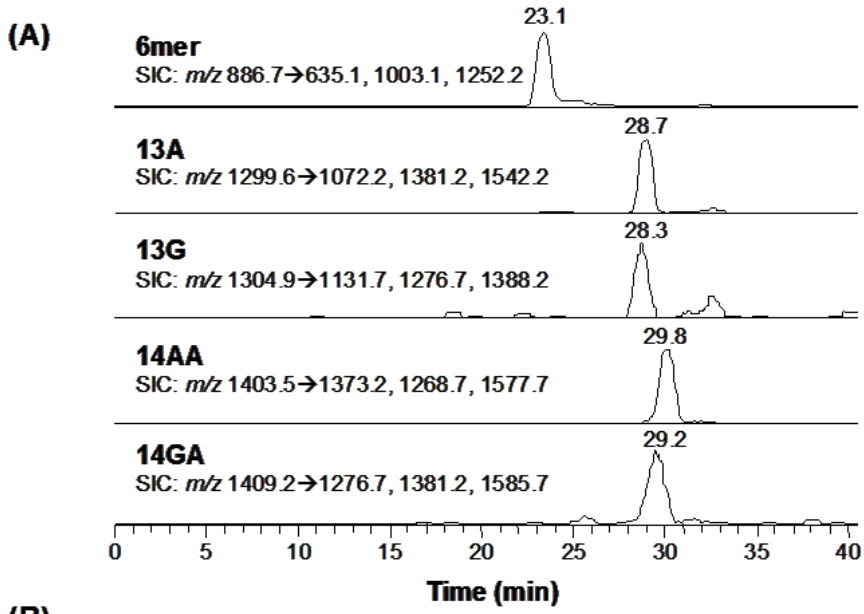
In-vitro replication with hPol η

The LC-MS/MS quantification results revealed that hPol η is capable of extending the primer to yield full-length replication products for all four substrates. In this vein, the amounts of unextended primer (6mer) represent 25%, 23%, 27% and 26% of the total replication products for substrates containing dT, O^2-EtdT, *N*3-EtdT and O^4-EtdT, respectively. This finding parallels the aforementioned PAGE result showing that the three EtdT lesions are not strong blocks to DNA replication catalyzed by hPol η (Figure 3.2A). In addition, we observed significant amount of blunt-end elongation products where a dAMP was added to the 3' terminus of the fully extended primer (Table 3.1; Figures 3.3 and B.7 and B.8).

Table 3.1. A summary of the percentages of products formed from hPol η-mediated replication of dT-, O^2-EtdT-, $N3$-EtdT- and O^4-EtdT-containing substrates as determined by LC-MS/MS. The bases highlighted in bold represent those incorporated during primer extension. Listed under sequences are the SacI cleavage products bearing the extended portion of the primer.

Name	Sequence	dT	O^2-EtdT	$N3$-EtdT	O^4-EtdT
6mer unextended primer	5'-CATAGC-3'	25	23	27	26
13A +7	5'-CATAGC**ACGCCAT**-3'	49	12		2
13G +7	5'-CATAGC**GCGCCAT**-3'		5		4
14AA (blunt end elongation) +8	5'-CATAGC**ACGCCATA**-3'	25	41	42	17
14GA (blunt end elongation) +8	5'-CATAGC**GCGCCATA**-3'		19	31	51

Figure 3.3. Selected-ion chromatograms obtained from the LC-MS and MS/MS analysis of the hPol η-induced replication products for O^2-EtdT-bearing substrate. The replication products were treated with restriction enzyme SacI and shrimp alkaline phosphatase, and 10-pmol of the replication mixture was injected for the analysis.



(B)

3′-CTAGCTCGAGTATCG**X**GCGGTA -5′		
6mer 5′-CATAGC -3′	23%	hPol η-induced Replication Products of O^2-EtdT Substrate
13A 5′-CATAGCACGCCAT -3′	12%	
13G 5′-CATAGCGCGCCAT -3′	5%	
14AA 5′-CATAGCACGCCATA-3′	41%	
14AG 5′-CATAGCGCGCCATA-3′	19%	

As expected, no mutagenic product was detectable for the control dT-housing substrate; nevertheless, all three EtdT lesions directed substantial frequencies of nucleotide misincorporation. The full-length products with dAMP and dGMP being inserted opposite the lesion site constitute 53% and 24% of all the products detected in the replication mixture for the O^2-EtdT-bearing substrate, 42% and 31% for the $N3$-EtdT-containing substrate, and 19% and 55% for the O^4-EtdT-carrying substrate (Table 3.1; Figures 3.3 and B.7 and B.8).

In-vitro replication with Kf

The three Et-dT lesions were moderately blocking to primer extension mediated by Kf˙, where the unextended primer and +1 nucleotide extension products together occupied 21%, 45%, 24% and 22% of all the detected replication products for substrates containing a dT, O^2-EtdT, $N3$-EtdT and O^4-EtdT, respectively (Table 3.2). Similar to hPol η, we again observed substantial amounts of blunt-end elongation products, where the additions of a dAMP, or both a dAMP and a dTMP, to the 3' termini of the full-length products were found (Table 3.2).

Table 3.2. A summary of the percentages of products formed from Kf$^-$-mediated replication of dT-, O^2-EtdT-, *N*3-EtdT- and O^4-EtdT-containing substrates as determined by LC-MS/MS. The bases highlighted in bold represent those incorporated during primer extension. Listed under sequences are the SacI cleavage products bearing the extended portion of the primer.

Name	Sequence	dT	O^2-EtdT	*N*3-EtdT	O^4-EtdT
6mer unextended primer	5'-CATAGC-3'	16	9	2	22
7T +1	5'-CATAGC**T**-3'			2	
7A +1	5'-CATAGC**A**-3'	5	36	12	
7G +1	5'-CATAGC**G**-3'			8	
8A +2	5'-CATAGC**AC**-3'	5			
13A +7	5'-CATAGC**ACGCCAT**-3'	10			
14TA (blunt end elongation) +8	5'-CATAGC**TCGCCATA**-3'		4	15	
14AA (blunt end elongation) +8	5'-CATAGC**ACGCCATA**-3'	11	21	17	17
14GA (blunt end elongation) +8	5'-CATAGC**GCGCCATA**-3'		4	21	35
15TAT (blunt end elongation) +9	5'-CATAGC**TCGCCATAT**-3'		3	7	
15AAT (blunt end elongation) +9	5'-CATAGC**ACGCCATAT**-3'	53	20	5	8
15GAT (blunt end elongation) +9	5'-CATAGC**GCGCCATAT**-3'		2	11	17

We also detected considerable frequencies of mutations for all three EtdT derivatives. Full-length replication products with the incorporation of dAMP, dGMP, and dTMP at the lesion site represent 41%, 6% and 7%, respectively, of all the detected replication products for O^2-EtdT-bearing substrate, while these percentages were 22%, 32% and 22% for the corresponding *N*3-EtdT substrate (Table 3.2). By contrast, we did not observe dTMP insertion opposite O^4-EtdT, and full-length replication products with dAMP and dGMP being incorporated opposite the lesion site contributed 25% and 52% to the total amounts of replication products, respectively (Table 3.2).

In-vitro replication with hPol κ

Human Pol κ was largely stalled by the three Et-dT lesions (particularly *N*3- and O^4-EtdT), where the unextended primer and the single-nucleotide extension products together represented 59%, 76% and 96% of the total amount of replication products for substrates containing O^2-, *N*3- and O^4-EtdT, respectively (Table 3.3). Although very small amounts (2%) of full-length products were found for substrates harboring an O^4-EtdT, full-length products with dAMP and dGMP being inserted opposite O^2-EtdT occurred at frequencies of 21% and 17%, respectively (Table 3.3). This finding is in reminiscent of the results shown in Figure 2D, where O^2-EtdT was found to be less blocking to hPol κ-mediated primer extension than O^4-EtdT and *N*3-EtdT. In addition, we found a relatively low frequency (3%, Table 3.3) of –1 deletion product for the O^2-EtdT substrate, where the polymerase skipped the lesion site during primer extension.

Table 3.3. A summary of the percentages of products formed from hPol κ-mediated replication of dT-, O^2-EtdT-, *N*3-EtdT- and O^4-EtdT-containing substrates as determined by LC-MS/MS. The bases highlighted in bold represent those incorporated during primer extension. Listed under sequences are the SacI cleavage products bearing the extended portion of the primer.

Name	Sequence	dT	O^2-EtdT	*N*3-EtdT	O^4-EtdT
6mer unextended primer	5'-CATAGC-3'	4	59	76	96
7A +1	5'-CATAGC**A**-3'				2
7G +1	5'-CATAGC**G**-3'				1
9C +3	5'-CATAGC**CCG**-3'			11	
9A +3	5'-CATAGC**ACG**-3'			1	
9G +3	5'-CATAGC**GCG**-3'			6	
10Del (at lesion site) +4	5'-CATAGC**CGCC**-3'			3	
10ADel (+1 from lesion site)+4	5'-CATAGC**AGCC**-3'			3	
11 Del (at lesion site) +5	5'-CATAGC**CGCCA**-3'		3		
12A +6	5'-CATAGC**ACGCCA**-3'	96	21		1
12G +6	5'-CATAGC**GCGCCA**-3'		17		1

In-vitro replication with hPol ι

Similar as what was revealed from PAGE analysis, LC-MS/MS results showed that the primer extension was highly inefficient, where hPol ι largely stalled after incorporation of a single nucleotide opposite the three Et-dT lesions. For O^2-EtdT, the insertion of dTMP (38%) dominated over those of dAMP (4%), dCMP (1%), and dGMP (5%, Table B.1). While the insertion of dGMP (64%) was favored over those of dAMP (2%) and dTMP (17%) for *N*3-EtdT, the insertion of the correct nucleotide, dAMP (28%), was preferred over those of the incorrect dCMP (6%), dGMP (16%) and dTMP (15%) for O^4-EtdT (Table B.1).

In-vitro replication with yPol ζ

O^2-EtdT and *N*3-EtdT were extremely strong blocks to yPol ζ, as manifested by dominance (93-96%) of the unextended primer in the replication mixture. This is again consistent with what we observed from PAGE analysis (Figure 3.2D). O^4-EtdT could be partially bypassed by yPol ζ, where the extended products predominantly carry the incorrect nucleotide (dGMP) opposite the lesion (Table B.2).

It is of note that the LC-MS results revealed that damage-containing substrate remains largely intact during the various stages of sample handling (i.e., primer extension and restriction digestion), where less than 3% of the O^2-EtdT-, *N*3-EtdT-, and O^4-EtdT-bearing templates were degraded to the corresponding unmodified dT-containing substrate (data not shown).

Discussion

Cigarette and its smoke contain thousands of compounds, some of which are known carcinogens. *In vitro* studies have revealed the formation of ethylated adducts in DNA exposed to cigarette smoke, though the identities of the direct-acting ethylating agents remain unknown.[6] In addition, several studies have demonstrated elevated levels of the ethylated DNA lesions in the tissue, blood and urine samples of smokers relative to non-smokers.[6-13, 42] In particular, a very recent study revealed that O^2-, $N3$-, and O^4-EtdT could be detected at higher levels in leukocyte DNA of smokers than non-smokers.[13] Despite a substantial amount of work conducted to assess the formation of these DNA lesions, no systematic study has been carried out to interrogate how different the three regioisomeric EtdT lesions are recognized by DNA polymerases.

In this study, we employed denaturing PAGE and LC-MS/MS to monitor the primer extension products of the three EtdT-containing substrates mediated by five different DNA polymerases. We found that hPol η and Kf⁻ were minimally blocked by the EtdT lesions, and full-length products could be readily detected for all three lesions. With the exception of O^2-EtdT substrate for hPol κ and O^4-EtdT substrate for yPol ζ, the EtdT lesions were strong blocks to hPol κ, hPol ι and yPol ζ.

Our LC-MS/MS quantification results also underscored significant miscoding potentials of the three EtdT lesions. As observed previously for O^4-alkylated thymidine derivatives,[14, 15, 21] we found that O^4-EtdT instructed the polymerases to misincorporate preferentially dGMP. Interestingly, we observed that both O^2- and $N3$-EtdT could also direct substantial frequencies of dGMP misincorporation with multiple polymerases

(Tables 3.1-3.3 and B.1-B.2). Additionally, the presence of these two lesions in the template DNA results in appreciable levels of misincorporation of dTMP opposite the lesions, though this depends on the nature of the polymerases involved. Exposure of rats to a known ethylating agent, *N*-ethyl-*N*-nitrosourea (ENU), was shown to give rise to a specific AT→TA point mutation which resulted in the activation of the neu oncogene in neuroblastoma cells.[43] Additionally, treatment of Chinese hamster ovary cells with ENU led to GC→AT, AT→TA and AT→GC substitutions in the coding region of the *Hprt* gene,[44] where the GC→AT and AT→GC mutations were attributed to replicative bypass of O^6-EtdG and O^4-EtdT, respectively.[44] Our results suggest that O^2-EtdT and *N*3-EtdT may also lead to substantial frequencies of AT→GC and AT→TA mutations. In this respect, some previous studies suggested that *N*3-EtdT and O^2-EtdT are capable of pairing with thymine during DNA synthesis.[18-20]

The better tolerance of hPol κ toward O^2-EtdT than the other two regioisomers is in line with the previous observation that the polymerase can carry out efficient and error-free nucleotide incorporation opposite the minor-groove N^2-dG lesions.[45-50] The poor fidelity and, relative to the bypass of the N^2-dG lesions, the reduced efficiency, observed for hPol κ-mediated nucleotide incorporation opposite O^2-EtdT might be attributed to the inability of O^2-EtdT to base pair favorably with any canonical nucleotides, as noted previously.[21] Taken together, this study reveals the biological consequences of the regioisomeric EtdT lesions by offering important knowledge about how these lesions compromise the efficiency and fidelity of DNA replication mediated by a number of DNA polymerases.

References

(1) Society, A. C. (2013) *Cancer Facts and Figures 2013*. American Cancer Society, Atlanta, GA.

(2) (2004) Tobacco smoke and involuntary smoking. *IARC Monographs Evaluation of Carcinogen Risk to Humans*, *83*, 1-1438.

(3) Tyczynski, J. E., Bray, F., and Parkin, D. M. (2003) Lung cancer in Europe in 2000: epidemiology, prevention, and early detection. *Lancet Oncol.*, *4*, 45-55.

(4) Hecht, S. S. (2008) Progress and challenges in selected areas of tobacco carcinogenesis. *Chem. Res. Toxicol.*, *21*, 160-171.

(5) Singer, B. (1985) *In vivo* formation and persistence of modified nucleosides resulting from alkylating agents. *Environ. Health Perspect.*, *62*, 41-48.

(6) Singh, R., Kaur, B., and Farmer, P. B. (2005) Detection of DNA damage derived from a direct acting ethylating agent present in cigarette smoke by use of liquid chromatography-tandem mass spectrometry. *Chem. Res. Toxicol.*, *18*, 249-256.

(7) Carmella, S. G., Chen, M., Villalta, P. W., Gurney, J. G., Hatsukami, D. K., and Hecht, S. S. (2002) Ethylation and methylation of hemoglobin in smokers and non-smokers. *Carcinogenesis*, *23*, 1903-1910.

(8) Prevost, V., and Shuker, D. E. (1996) Cigarette smoking and urinary 3-alkyladenine excretion in man. *Chem. Res. Toxicol.*, *9*, 439-444.

(9) Chao, M. R., Wang, C. J., Chang, L. W., and Hu, C. W. (2006) Quantitative determination of urinary N7-ethylguanine in smokers and non-smokers using an isotope dilution liquid chromatography/tandem mass spectrometry with on-line analyte enrichment. *Carcinogenesis*, *27*, 146-151.

(10) Anna, L., Kovacs, K., Gyorffy, E., Schoket, B., and Nair, J. (2011) Smoking-related O^4-ethylthymidine formation in human lung tissue and comparisons with bulky DNA adducts. *Mutagenesis*, *26*, 523-527.

(11) Bhanot, O. S., Grevatt, P. C., Donahue, J. M., Gabrielides, C. N., and Solomon, J. J. (1992) *In vitro* DNA replication implicates O^2-ethyldeoxythymidine in transversion mutagenesis by ethylating agents. *Nucleic Acids Res.*, *20*, 587-594.

(12) Balbo, S., Villalta, P. W., and Hecht, S. S. (2011) Quantitation of 7-ethylguanine in leukocyte DNA from smokers and nonsmokers by liquid chromatography-nanoelectrospray-high resolution tandem mass spectrometry. *Chem. Res. Toxicol.*, *24*, 1729-1734.

(13) Chen, H. J., Wang, Y. C., and Lin, W. P. (2012) Analysis of ethylated thymidine adducts in human leukocyte DNA by stable isotope dilution nanoflow liquid chromatography-nanospray ionization tandem mass spectrometry. *Anal. Chem.*, *84*, 2521-2527.

(14) Preston, B. D., Singer, B., and Loeb, L. A. (1986) Mutagenic potential of O^4-methylthymine *in vivo* determined by an enzymatic approach to site-specific mutagenesis. *Proc. Natl. Acad. Sci. U S A*, *83*, 8501-8505.

(15) Singer, B., Spengler, S. J., Fraenkel-Conrat, H., and Kusmierek, J. T. (1986) O^4-Methyl, -ethyl, or -isopropyl substituents on thymidine in poly(dA-dT) all lead to transitions upon replication. *Proc. Natl. Acad. Sci. U S A*, *83*, 28-32.

(16) Jasti, V. P., Spratt, T. E., and Basu, A. K. (2011) Tobacco-specific nitrosamine-derived O^2-alkylthymidines are potent mutagenic lesions in SOS-induced *Escherichia coli*. *Chem. Res. Toxicol.*, *24*, 1833-1835.

(17) Gowda, A. S., Krishnegowda, G., Suo, Z., Amin, S., and Spratt, T. E. (2012) Low fidelity bypass of O^2-(3-pyridyl)-4-oxobutylthymine, the most persistent bulky adduct produced by the tobacco specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone by model DNA polymerases. *Chem. Res. Toxicol.*, *25*, 1195-1202.

(18) Fiala, K. A., Brown, J. A., Ling, H., Kshetry, A. K., Zhang, J., Taylor, J. S., Yang, W., and Suo, Z. (2007) Mechanism of template-independent nucleotide incorporation catalyzed by a template-dependent DNA polymerase. *J. Mol. Biol.*, *365*, 590-602.

(19) Grevatt, P. C., Solomon, J.J. and Bhanot, O.S. (1992) *In vitro* mispairing specificity of O^2-ethylthymidine. *Biochemistry*, *31*, 4181-4188.

(20) Huff, A. C., and Topal, M. D. (1987) DNA damage at thymine *N*3 abolishes base-pairing capacity during DNA-synthesis. *J. Biol. Chem.*, *262*, 12843-12850.

(21) Andersen, N., Wang, J., Wang, P., Jiang, Y., and Wang, Y. (2012) *In-vitro* replication studies on O^2-methylthymidine and O^4-methylthymidine. *Chem. Res. Toxicol.*, *25*, 2523-2531.

(22) Bronstein, S. M., Skopek, T. R., and Swenberg, J. A. (1992) Efficient repair of O^6-ethylguanine, but not O^4-ethylthymine or O^2-ethylthymine, is dependent upon O^6-alkylguanine-DNA alkyltransferase and nucleotide excision repair activities in human cells. *Cancer Res.*, *52*, 2008-2011.

(23) Brent, T. P., Dolan, M. E., Fraenkel-Conrat, H., Hall, J., Karran, P., Laval, L., Margison, G. P., Montesano, R., Pegg, A. E., Potter, P. M., and et al. (1988) Repair of *O*-alkylpyrimidines in mammalian cells: a present consensus. *Proc. Natl. Acad. Sci. U S A*, *85*, 1759-1762.

(24) Den Engelse, L., De Graaf, A., De Brij, R. J., and Menkveld, G. J. (1987) O^2- and O^4-ethylthymine and the ethylphosphotriester dTp(Et)dT are highly persistent DNA modifications in slowly dividing tissues of the ethylnitrosourea-treated rat. *Carcinogenesis*, *8*, 751-757.

(25) Tissier, A., McDonald, J. P., Frank, E. G., and Woodgate, R. (2000) polι, a remarkably error-prone human DNA polymerase. *Genes Dev.*, *14*, 1642-1650.

(26) Biertumpfel, C., Zhao, Y., Kondo, Y., Ramon-Maiques, S., Gregory, M., Lee, J. Y., Masutani, C., Lehmann, A. R., Hanaoka, F., and Yang, W. (2010) Structure and mechanism of human DNA polymerase η. *Nature*, *465*, 1044-1048.

(27) Wang, J., and Wang, Y. (2009) Chemical synthesis of oligodeoxyribonucleotides containing *N*3- and O^4-carboxymethylthymidine and their formation in DNA. *Nucleic Acids Res.*, *37*, 336-345.

(28) Xu, Y. Z., and Swann, P. F. (1990) A simple method for the solid phase synthesis of oligodeoxynucleotides containing O^4-alkylthymine. *Nucleic Acids Res.*, *18*, 4061-4065.

(29) Xu, Y. Z., and Swann, P. F. (1994) Oligodeoxynucleotides containing O^2-alkylthymine: synthesis and characterization. *Tetrahedron Lett.*, *35*, 303-306.

(30) Yuan, B., Wang, J., Cao, H., Sun, R., and Wang, Y. (2011) High-throughput analysis of the mutagenic and cytotoxic properties of DNA lesions by next-generation sequencing. *Nucleic Acids Res*, *39*, 5945-5954.
(31) Choi, J. Y., Zang, H., Angel, K. C., Kozekov, I. D., Goodenough, A. K., Rizzo, C. J., and Guengerich, F. P. (2006) Translesion synthesis across 1,N^2-ethenoguanine by human DNA polymerases. *Chem. Res. Toxicol.*, *19*, 879-886.
(32) Kotapati, S., Maddukuri, L., Wickramaratne, S., Seneviratne, U., Goggin, M., Pence, M. G., Villalta, P., Guengerich, F. P., Marnett, L., and Tretyakova, N. (2012) Translesion synthesis across 1,N^6-(2-hydroxy-3-hydroxymethylpropan-1,3-diyl)-2'-deoxyadenosine (1,N^6-γ-HMHP-dA) adducts by human and archebacterial DNA polymerases. *J. Biol. Chem.*, *287*, 38800-38811.
(33) Washington, M. T., Johnson, R. E., Prakash, L., and Prakash, S. (2001) Accuracy of lesion bypass by yeast and human DNA polymerase η. *Proc. Natl. Acad. Sci. U. S. A.*, *98*, 8355-8360.
(34) Nelson, J. R., Lawrence, C. W., and Hinkle, D. C. (1996) Thymine-thymine dimer bypass by yeast DNA polymerase ζ. *Science*, *272*, 1646-1649.
(35) Pence, M. G., Blans, P., Zink, C. N., Hollis, T., Fishbein, J. C., and Perrino, F. W. (2009) Lesion bypass of N^2-ethylguanine by human DNA polymerase ι. *J. Biol. Chem.*, *284*, 1732-1740.
(36) Choi, J. Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L. A., and Guengerich, F. P. (2006) Translesion synthesis across O^6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J. Biol. Chem.*, *281*, 38244-38256.
(37) Frank, E. G., Tissier, A., McDonald, J. P., Rapic-Otrin, V., Zeng, X., Gearhart, P. J., and Woodgate, R. (2001) Altered nucleotide misinsertion fidelity associated with polι-dependent replication at the end of a DNA template. *EMBO J.*, *20*, 2914-2922.
(38) Vaisman, A., Takasawa, K., Iwai, S., and Woodgate, R. (2006) DNA polymerase ι-dependent translesion replication of uracil containing cyclobutane pyrimidine dimers. *DNA Repair (Amst.)*, *5*, 210-218.
(39) Vaisman, A., and Woodgate, R. (2001) Unique misinsertion specificity of polι may decrease the mutagenic potential of deaminated cytosines. *EMBO J.*, *20*, 6520-6529.
(40) Chowdhury, G., and Guengerich, F. P. (2011) Liquid chromatography-mass spectrometry analysis of DNA polymerase reaction products. *Curr. Protoc. Nucleic Acid Chem.*, *Chapter 7*, Unit 7 16 11-11.
(41) Gu, C., and Wang, Y. (2007) *In vitro* replication and thermodynamic studies of methylation and oxidation modifications of 6-thioguanine. *Nucleic Acids Res.*, *35*, 3693-3704.
(42) Godschalk, R., Nair, J., van Schooten, F. J., Risch, A., Drings, P., Kayser, K., Dienemann, H., and Bartsch, H. (2002) Comparison of multiple DNA adduct types in tumor adjacent human lung tissue: effect of cigarette smoking. *Carcinogenesis*, *23*, 2081-2086.

(43) Perantoni, A. O., Rice, J. M., Reed, C. D., Watatani, M., and Wenk, M. L. (1987) Activated neu oncogene sequences in primary tumors of the peripheral nervous system induced in rats by transplacental exposure to ethylnitrosourea. *Proc. Natl. Acad. Sci. U. S. A.*, *84*, 6317-6321.

(44) Op het Veld, C. W., van Hees-Stuivenberg, S., van Zeeland, A. A., and Jansen, J. G. (1997) Effect of nucleotide excision repair on hprt gene mutations in rodent cells exposed to DNA ethylating agents. *Mutagenesis*, *12*, 417-424.

(45) Jarosz, D. F., Godoy, V. G., Delaney, J. C., Essigmann, J. M., and Walker, G. C. (2006) A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature*, *439*, 225-228.

(46) Yuan, B., Cao, H., Jiang, Y., Hong, H., and Wang, Y. (2008) Efficient and accurate bypass of N^2-(1-carboxyethyl)-2'-deoxyguanosine by DinB DNA polymerase *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. USA*, *105*, 8679-8684.

(47) Yuan, B., You, C., Andersen, N., Jiang, Y., Moriya, M., O'Connor, T. R., and Wang, Y. (2011) The roles of DNA polymerases κ and ι in the error-free bypass of N^2-carboxyalkyl-dG lesions in mammalian cells. *J Biol Chem*, *286*, 17503-17511.

(48) Choi, J. Y., Angel, K. C., and Guengerich, F. P. (2006) Translesion synthesis across bulky N^2-alkyl guanine DNA adducts by human DNA polymerase κ. *J. Biol. Chem.*, *281*, 21062-21072.

(49) Suzuki, N., Ohashi, E., Kolbanovskiy, A., Geacintov, N. E., Grollman, A. P., Ohmori, H., and Shibutani, S. (2002) Translesion synthesis by human DNA polymerase κ on a DNA template containing a single stereoisomer of dG-(+)- or dG-(-)-*anti*-N^2-BPDE (7,8-dihydroxy-*anti*-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene). *Biochemistry*, *41*, 6100-6106.

(50) Zhang, Y., Yuan, F., Wu, X., Wang, M., Rechkoblit, O., Taylor, J. S., Geacintov, N. E., and Wang, Z. (2000) Error-free and error-prone lesion bypass by human DNA polymerase κ *in vitro*. *Nucleic Acids Res.*, *28*, 4138-4146.

Chapter 4: The Roles of DNA Polymerases κ and ι in the Error-free Bypass of N^2-Carboxyalkyl-2′-deoxyguanosine Lesions in Mammalian Cells

Introduction

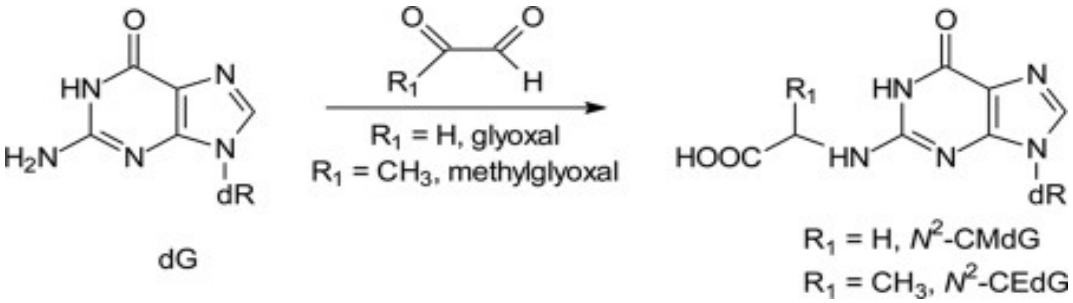
DNA is susceptible to damage by endogenous and exogenous agents.[1] To minimize cell death arising from replication blockage and mutations emanating from nucleotide misincorporation opposite the damage site, cells are equipped with several pathways to repair DNA lesions.[2] Additionally, cells have evolved two major mechanisms to tolerate unrepaired DNA lesions,[3] including homologous recombination and translesion synthesis (TLS).[3] In this regard, synthesis past many DNA lesions requires the replacement of replicative DNA polymerases with one or a few specialized polymerases, most of which belong to the Y-family.[4] Along this line, phylogenetic analysis revealed five branches of Y-family polymerases, which include UmuC, DinB, polymerase η, polymerase ι, and REV1.[4, 5] All of these, except for UmuC, are found in eukaryotic cells.[4, 5] Apart from the Y-family DNA polymerases, some B-family DNA polymerases also participate in TLS. These include DNA polymerase II in *Escherichia coli* and polymerase ζ, which is composed of the catalytic subunit REV3 and the regulatory subunit REV7, in eukaryotic cells.[4, 5]

These specialized DNA polymerases are characterized by their relatively low fidelity in replicating undamaged DNA but have the capability to bypass those DNA lesions that

normally block DNA synthesis by replicative DNA polymerases.[4] It was advocated that specialized DNA polymerases are evolved by nature to copy DNA base damage accurately, but they no longer bear the ability to replicate unmodified DNA with high fidelity.[6] The best known example of this is the efficient and accurate bypass of *cis*,*syn*-cyclobutane pyrimidine dimers by polymerase η,[7, 8] which is encoded by the xeroderma pigmentosum-variant gene (*i.e. POLH*) in humans.[9] Xeroderma pigmentosum-variant patients exhibit elevated susceptibility to developing skin cancer.[10] Several recent studies also demonstrated that DinB DNA polymerase, a Y-family polymerase conserved in all three kingdoms of life,[11] can insert the correct dCMP opposite several N^2-substituted guanine lesions at an efficiency that is similar to or better than opposite an unmodified guanine.[12-14] The x-ray crystal structure of the catalytic core of human DNA polymerase κ, along with primer/template DNA and an incoming nucleotide, reveals the lack of steric hindrance in the minor groove at the primer-template junction,[15] which may account for the tolerance of polymerase κ toward the minor groove N^2-dG lesions. By using shuttle vector technology, we also showed that DinB (*i.e.* polymerase IV) is the major polymerase involved in the accurate bypass of the two diastereomers of N^2-(1-carboxyethyl)-2′-deoxyguanosine (N^2-CEdG) in *E. coli* cells.[14] However, it remains unclear whether this finding can be extended to mammalian cells and whether other mammalian TLS polymerase(s) are also required for the error-free bypass of N^2-CEdG.

N^2-CEdG is the major stable DNA adduct formed from methylglyoxal, a glycolytic metabolite (Figure 4.1).[16] This modified nucleoside was found to be present in urine

Figure 4.1. The formation of N^2-CedG and N^2-CMdG.



samples from healthy human subjects,[17] and the lesion was observed more frequently in kidney and aortic cells of diabetic and uremic patients than in the corresponding cells of healthy human subjects.[18] N^2-CEdG can also be detected in cultured human cells, and the exposure of these cells to methylglyoxal or glucose further enhanced the formation of this lesion.[14] Moreover, recent studies illustrated that the exposure of isolated DNA to the structurally related glyoxal can give rise to the formation of N^2-carboxymethyl-2′-deoxyguanosine (N^2-CMdG).[19, 20] The major objective of the present study is to define the roles of polymerase κ and other mammalian TLS DNA polymerases in bypassing the N^2-CEdG and N^2-CMdG lesions in cells.

Materials and Methods

Materials

Unmodified ODNs used in this study were purchased from Integrated DNA Technologies (Coralville, IA). [γ-^{32}P]ATP was obtained from Perkin Elmer Life Sciences, and chemicals unless otherwise noted were from Sigma-Aldrich. Human polymerases η and ι were purchased from Enzymax (Lexington, KY). Shrimp alkaline phosphatase was obtained from USB Corp. (Cleveland, OH), and all other enzymes unless otherwise specified were from New England Biolabs.

The *Polk*-, *Rev3*-, and *Poli*-deficient mouse embryonic fibroblast (MEF) cells were kindly provided by Professors Haruo Ohmori,[21] Richard D. Wood,[22] and Roger Woodgate, [23] respectively. The C-Tag cells (*POLH*-deficient cells), which are SV40-immortalized cells derived from xeroderma pigmentosum-variant patients, were provided generously by Professor William K. Kaufmann.[24]

Preparation and LC-MS/MS Characterizations of ODN Substrates Containing an (S)-N^2-CEdG, (R)-N^2-CEdG, or N^2-CMdG

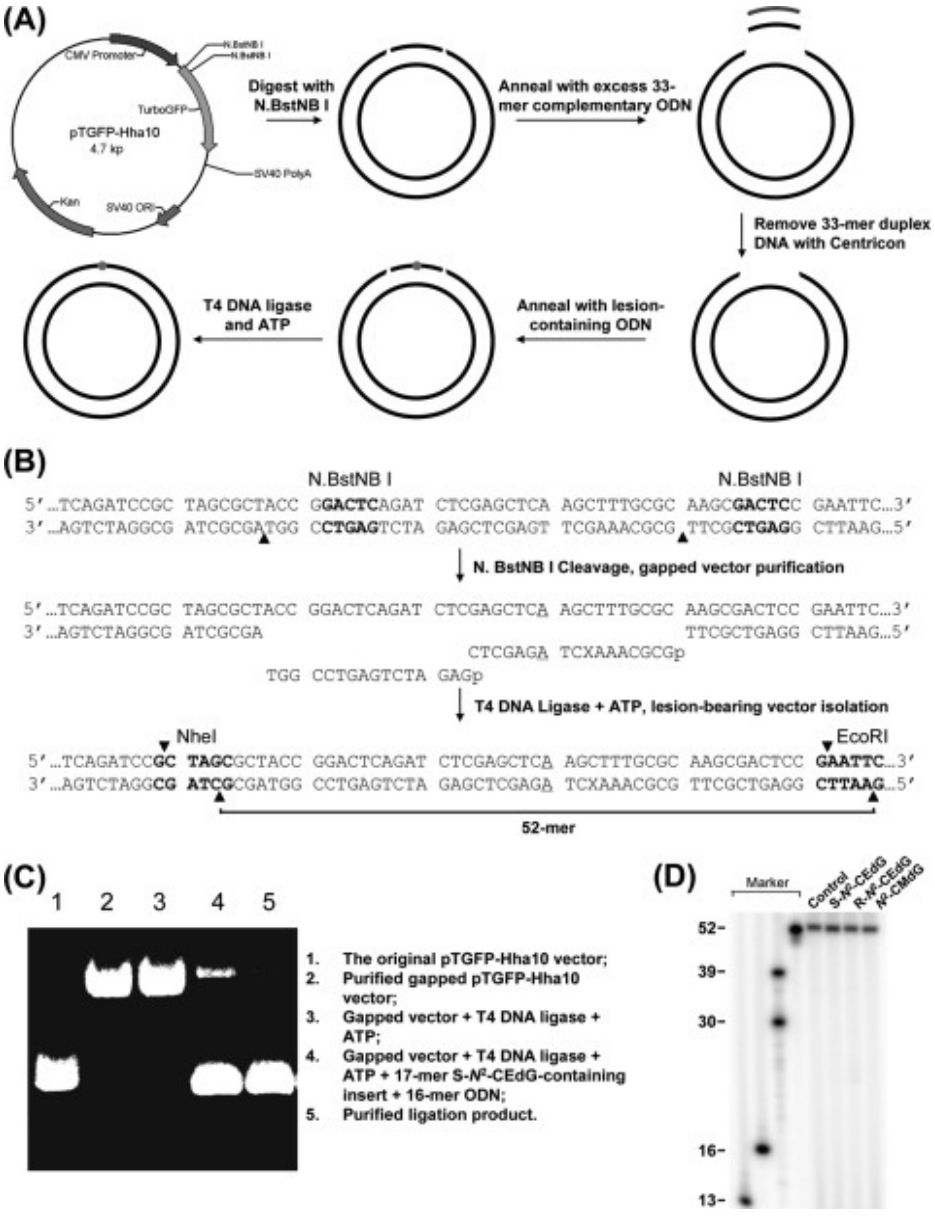
The 17-mer ODN 5′-GCGCAAA*X*CTAGAGCTC-3′ (*X* = N^2-CEdG) was synthesized following previously reported procedures.[25] The corresponding ODN bearing an N^2-CMdG was prepared by using the same procedures except that glycine was used in lieu of alanine to displace the fluorine atom in 2-fluoro-2′-deoxyinosine after ODN assembly.

The identities of N^2-CEdG- and N^2-CMdG-containing 17-mer ODNs were confirmed by online capillary HPLC electrospray ionization-MS/MS (Figures C.1 and C.2) using an Agilent 1100 capillary HPLC pump (Agilent Technologies) interfaced with an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA), which was configured for monitoring the fragmentation of the $[M - 4H]^{4-}$ ions of the 17-mer ODNs. A 5-min linear gradient of 0–30% methanol followed by a 30 min of 30–50% methanol in 400 mm 1,1,1,3,3,3-hexafluoro-2-propanol buffer (pH adjusted to 7.0 by addition of triethylamine) was employed.

Construction of pTGFP-Hha10 and pMTEX4 Genomes Harboring a Site-specifically Inserted (S)-N^2-CEdG, (R)-N^2-CEdG, N^2-CMdG, or dG

A previously described method[26] was used to insert a structurally defined lesion into a unique site in the pTGFP-Hha10 vector. Briefly, we first nicked the original vector with N.BstNBI to produce a gapped vector and a 33-mer single-stranded DNA (see Figures 4.2A and B). In this respect, the pTGFP-Hha10 vector contains two N.BstNBI recognition sites upstream of the coding region of the Turbo GFP (TGFP) gene, and

Figure 4.2. Construction and characterization of the lesion-bearing pTGFP-Hha10 shuttle vector. A schematic diagram showing the procedures for the preparation of the lesion-containing plasmid (A). Enzymatic digestion and ligation for the insertion of damage-carrying ODN into the gapped pTGFP-Hha10 vector (B). Agarose (1%) gel electrophoresis for monitoring the processes for the construction of the (*S*)-N^2-CEdG-bearing pTGFP-Hha10 vector (C). Restriction digestion and post-labeling for assessing the integrity of the control and lesion-containing vectors (D).



N.BstNBI nicks only one strand of duplex DNA at the fourth nucleotide 3′ to the GAGTC restriction recognition site. The 33-mer single-stranded ODN was removed subsequently from the nicked plasmid by annealing the cleavage mixture with the complementary 33-mer ODN in 50-fold molar excess. The gapped plasmid was isolated from the mixture by using 100-kDa cut-off ultracentrifugation filter units (Centricon 100 from Millipore). The purified gapped construct (see Figure 4.2C, *lane 2*) was subsequently annealed with a 5′-phosphorylated 17-mer lesion-carrying insert and a 16-mer unmodified ODN and ligated with T4 DNA ligase (see Figure 4.2B). In this regard, we used a 17-mer rather than 33-mer lesion-containing insert because of the relative ease in the synthesis and purification of the former substrate. The ligation mixture was incubated with ethidium bromide, and the resulting supercoiled, lesion-carrying double-stranded plasmid was isolated from the mixture by agarose gel electrophoresis (see Figure 4.2C, *lanes 4* and *5*). The pTGFP-Hha10 vector contains an SV40 origin, which allows the vectors to be replicated in the SV40-transformed human 293T and *POLH*-deficient cells.

The replicating vector pMTEX4 was used for experiments conducted in the MEF cell lines. pMTEX4 contains a polyomavirus T antigen gene and a polyomavirus origin of replication, which allow pMTEX4 to be replicated in MEF cells. A 54-bp duplex DNA with Mlu I and NheI cohesive ends was ligated to the Mul I-NheI digested pMTEX4. The resulting pMTEX4 vector was nicked with Nb.BbvCI to produce a gapped vector and a 41-mer single-stranded DNA. The subsequent procedures were similar as those for the construction of lesion-containing pTGFP-Hha10 vector as described above.

Cell Culture, Transfection, and in Vivo Replication

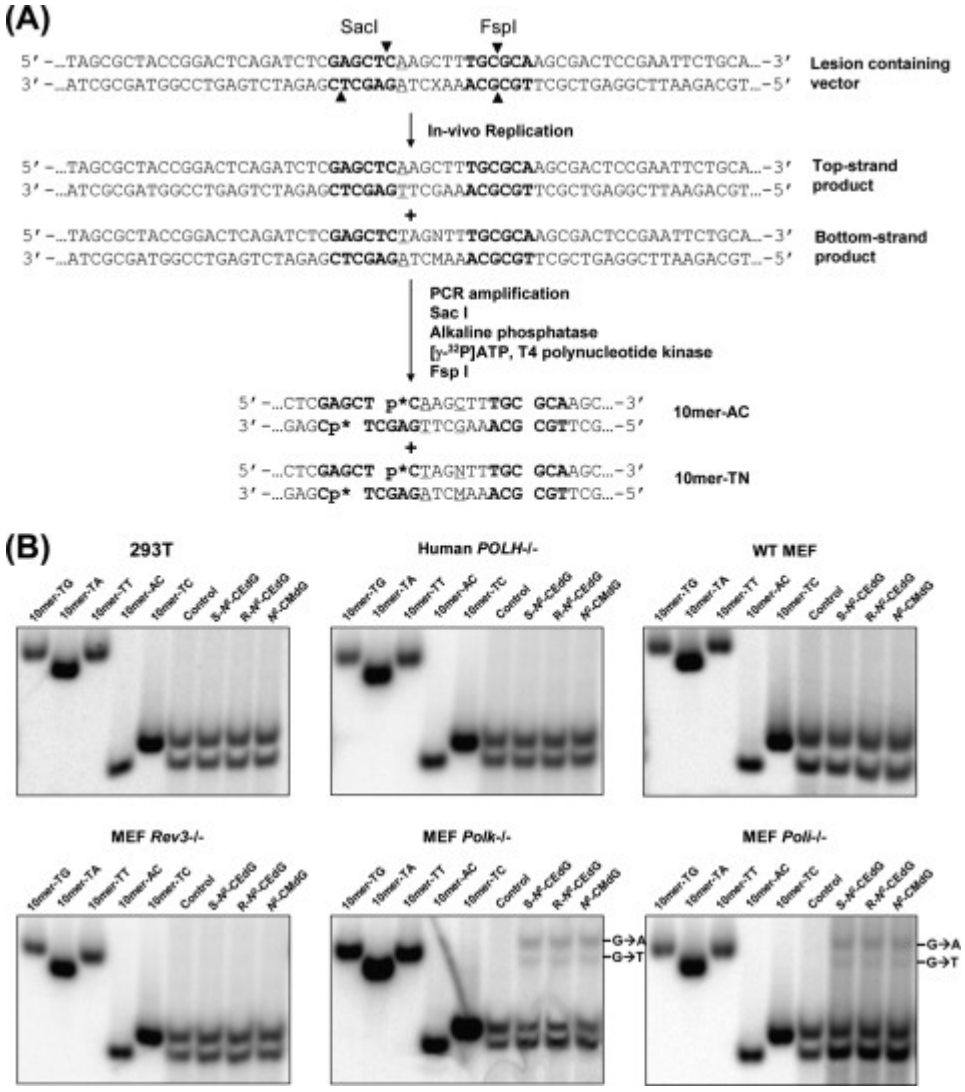
The 293T, *POLH*-deficient human fibroblast cells and all MEF cells were cultured at 37 °C in a 5% CO_2 atmosphere and in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (Invitrogen), 100 units/ml penicillin, and 100 μg/ml streptomycin (ATCC, Manassas, VA). Cells (5×10^5) were seeded in six-well plates and cultured overnight, after which they were transfected with 300 ng of the lesion-containing or lesion-free pTGFP-Hha10 or pMTEX4 vector using Lipofectamine 2000 (Invitrogen). After *in vivo* replication for 24 h, DNase I was added to digest the residual DNA adsorbed to the exterior of cells.[27] Cells were subsequently detached by treating with trypsin-EDTA (ATCC), the progenies of the plasmid were isolated by using an alkali lysis method,[28] and residual unreplicated plasmid was further removed by DpnI digestion.

Determination of Bypass Efficiency and Mutation Frequency

The bypass efficiencies and mutation frequencies were determined by a method adapted from the restriction endonuclease and post-labeling assay,[29, 30] which was developed initially by Essigmann and co-workers[31-33] for assessing the cytotoxic and mutagenic properties of DNA lesions in *E. coli* cells using the single-stranded M13 genome. The progeny genomes were amplified subsequently by PCR using Phusion high fidelity DNA polymerase. The primers were 5′-GCAGAGCTGGTTTAGTGAACCGTCAG-3′ and 5′-CTCTGCTGAAGCCAGTTACCTTCGG-3′ for pTGFP-Hha10 vector, 5′-GCTATCGAATTAATACGACTCATTATAGG-3′ and 5′-CCTTCGGAAAAAGAGTTGGTAGC-3′ for pMTEX4 vector. Target DNA was

amplified for 35 cycles, each consisting of 10 s at 98 °C, 30 s at 62 °C, 2 min at 72 °C, with a final extension at 72 °C for 5 min. The resulting 3950-bp PCR products (from pTGFP-Hha10 progenies) and 3412-bp PCR products (from pMTEX4 progenies) were purified by using a QIAquick PCR purification kit (Qiagen, Valencia, CA). For the bypass efficiency assay, a portion of the above PCR fragments was treated with 10 units of SacI and 1 unit of shrimp alkaline phosphatase (USB Corp.) in 10-μl New England Biolabs buffer 4 at 37 °C for 1 h, followed by heating at 65 °C for 20 min to deactivate the phosphatase. The above mixture was then treated in a 15-μl NEB buffer 4 with 5 mm DTT, ATP (50 pmol (cold), premixed with 1.66 pmol [γ-^{32}P]ATP), and 10 units of polynucleotide kinase. The reaction was continued at 37 °C for 1 h, followed by heating at 65 °C for 20 min to deactivate the polynucleotide kinase. 10 units of FspI was added to the reaction mixture (see recognition sites highlighted in *boldface* in Figure 3-3A), and the solution was incubated at 37 °C for 1 h, followed by quenching with 15 μl of formamide gel loading buffer containing xylene cyanol FF and bromphenol blue dyes. The mixture was loaded onto a 30% native polyacrylamide gel (acrylamide:bis-acrylamide = 19:1), and products were quantified by a phosphorimaging analysis.[14] After these cleavages, the original lesion site was housed in a 14-mer/10-mer duplex, d(pGCAAAMCTAGAGCT)/d(p*CTAGNTTTGC) (bottom strand product), where M represents the nucleobase incorporated at the initial damage site during *in vivo* DNA replication, N is the paired nucleobase of M in the complementary strand, and p* designates the 5′-radiolabeled phosphate (see Figure 4.3A). The mutation frequencies were determined from the relative amounts of different 10-mer products from the gel

Figure 4.3. Determination of bypass efficiencies and mutation frequencies of (*S*)- or (*R*)-N^2-CEdG and N^2-CMdG. Restriction digestion of PCR products of the progeny genome arising from *in vivo* replication of N^2-CEdG- and N^2-CMdG-bearing genome in mammalian cells (A). Nondenaturing PAGE for assessing the bypass efficiencies and mutation frequencies of N^2-CEdG and N^2-CMdG in mammalian cells that are proficient in TLS DNA polymerase (human 293T cells and wild-type MEF cells) or deficient in polymerase η (*POLH*$^{-/-}$), polymerase ζ (*Rev3*$^{-/-}$), polymerase κ (*Polk*$^{-/-}$), and polymerase ι (*Poli*$^{-/-}$) (B).



band intensities. On the other hand, the restriction cleavage of the product arising from the replication of the lesion-free top strand gave d(pGCAAAGCTTGAGCT)/d(p*CAAGCTTTGC). The 10-mer products were monitored instead of the 14-mer products because the former products could be resolved by native polyacrylamide gel. The bypass efficiency was calculated using the following formula, % bypass = (lesion vector bottom strand signal/lesion vector top strand signal)/(control vector bottom strand signal/control vector top strand signal).

Identification of Replication Products by Using LC-MS/MS

To identify the replication products using LC-MS, a nested PCR was performed using primers 5′-ACGGTGCCGACAGGTGCTTC-3′ and 5′-CCCTTCCGGAGACGACTAGTGC-3′. The 238-bp PCR products were treated with 50 units of SacI, 50 units of FspI, and 20 units of shrimp alkaline phosphatase in a 250-μl NEB buffer 4 at 37 °C for 2 h, followed by heating at 65 °C for 20 min. The resulting solution was extracted with phenol/chloroform/isoamyl alcohol (25:24:1, v/v), and the aqueous portion was dried with Speed-vac and dissolved in 12 μl of water. The ODN mixture was subjected to LC-MS/MS analysis. A 0.5 × 150 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies) was used for the separation, and the flow rate was 8.0 μl/min, which was delivered by using an Agilent 1100 capillary HPLC pump. A 5-min gradient of 0–20% methanol followed by a 35-min of 20–50% methanol in 400 mm 1,1,1,3,3,3-hexafluoro-2-propanol (pH was adjusted to 7.0 by the addition of triethylamine) was employed for the separation. The effluent from the LC column was coupled directly to an LTQ linear ion trap mass spectrometer (Thermo Electron, San

Jose, CA), which was configured for monitoring the fragmentation of the $[M - 3H]^{3-}$ ions of the 14-mer ODNs (*i.e.* d(GCAAAMCTAGAGCT), where M designates A, T, C, or G).

siRNA Knockdown of POLK Gene in 293T Cells and Replication Studies in siRNA-treated Cells

ON-TARGET*plus* SMARTpool (L-021038) and siGENOME Nontargeting siRNA (D-001210) were purchased from Dharmacon (Lafayette, CO). 293T cells were grown in DMEM medium (ATCC) containing 10% fetal bovine serum (Invitrogen) and penicillin/streptomycin (ATCC). Cells were seeded in six-well plates at 70% confluence and transfected with 1 μg of synthetic duplex siRNAs using Lipofectamine 2000 following the manufacturer's instructions. After a 48-h incubation, cells were co-transfected with another aliquot of siRNA and pTGFP-Hha10 lesion-containing vector followed by isolation of progeny vectors after a 24-h incubation with the same methods used previously. The isolated progeny vectors were processed using a protocol identical to that described above.

Real-time Quantitative RT-PCR and Immunoblot Analysis

Total RNA was extracted from the cells 48 and 72 h after transfection with siRNA using the RNeasy Mini Kit (Qiagen). cDNA was synthesized by using iScript^{TM} cDNA synthesis kit (Bio-Rad) according to the manufacturer's recommended procedures. Briefly, 1 μg of total RNA was reverse-transcribed with 1 μl of iScript reverse transcriptase in a 20-μl volume reaction. The reaction was carried out at 25 °C for 5 min and at 42 °C for 30 min. The reverse transcriptase was then deactivated by heating at 85 °C for 5 min. Analyses of the transcripts for *POLK* were performed by real-time quantitative RT-PCR using iQ^{TM} SYBR Green Supermix kit (Bio-Rad) according to the

manufacturer's recommendations on a Bio-Rad iCycler system (Bio-Rad). The following primers were used for the real-time PCR: 5′-TGCTGATTTTCCACATCCCTTGAG-3′ and 5′-CTCCTTTGTTGGTGTTTCCTGTCC-3′ for *POLK*, 5′-TTTGTCAAGCTCATTTCCTGGTATG-3′ and 5′-TCTCTTCCTCTTGTGCTCTTGCTG-3′ for GAPDH. The comparative cycle threshold (Ct) method ($\Delta\Delta Ct$) was used for the relative quantification of gene expression.[34] Immunoblot analysis was preformed with a total of 40 μg of whole cell lysate. Antibody that specifically recognizes human DNA polymerase κ (Santa Cruz Biotechnology) was used at a 1:2000 dilution. Human β-actin antibody (Abcam, Cambridge, MA) was used at a 1:5000 dilution. Horseradish peroxidase-conjugated secondary goat anti-mouse antibody (Santa Cruz Biotechnology) and goat anti-rabbit antibody (Abcam) were used at a 1:10,000 dilution.

Steady-state Kinetic Measurements

A 24-mer ODN, d(5′-GCGCAAAXCTAGAGCTCGAGATCT-3′) (X = S/(*R*)-N^2-CEdG) and a 5′-^{32}P-labeled 16-mer primer, d(AGATCTCGAGCTCTAG), were employed for steady-state kinetic measurements.

The 24-mer N^2-CEdG-containing ODNs, 5′-GCGCAAA*X*CTAGAGCTCGAGATCT-3′, *X* = (*S/R*)-N^2-CEdG), were prepared, and steady-state kinetic analyses were performed following procedures described previously.[14] The 24-mer lesion-containing template or nondamaged template (20 nm) with dG in place of the N^2-CEdG was annealed with a 5′-^{32}P-labeled 16-mer primer d(AGATCTCGAGCTCTAG) to give the primer-template complex. The primer-template complex (10 nm) was incubated at 37 °C with human polymerase η (5 ng) for 10 min or

polymerase ι (5 ng) for 30 min in the presence of various concentrations of an individual dNTP and in a buffer containing 20 mm Tris-HCl (pH 7.5), 2 mm dithioreitol, 100 μg/ml bovine seine albumin, 10% glycerol, and 5 mm (for polymerase η) or 2 mm (for polymerase ι) $MgCl_2$. The individual dNTP concentration was optimized for different insertion reactions to allow for a <20% primer extension.[35] The buffer contained 80% formamide, 10 mM EDTA (pH 8.0), 1 mg/mL xylene cyanol and 1 mg/mL bromophenol blue. The products were resolved on 20% (29:1) cross-linked polyacrylamide gels containing 8 M urea. Gel band intensities for the substrates and products were quantified using a Typhoon 9410 variable-mode imager (Amersham Biosciences Co.) and ImageQuant version 5.2 (Amersham Biosciences Co.). The dNTP concentration was optimized for different insertion reactions to allow for approximately 20% primer extension.[35] The observed rate of dNTP incorporation (V_{obs}) was plotted versus the dNTP concentration, and the apparent steady-state kinetic parameters (K_m and V_{max}) for the incorporation of the correct and incorrect nucleotides were determined by fitting the rate data with the Michaelis-Menten equation:

$$V_{obs} = \frac{V_{\max} \times [dNTP]}{K_m + [dNTP]}$$

The k_{cat} values were then calculated by dividing the V_{max} values with the concentration of the polymerase used. The efficiency of nucleotide incorporation was determined by the ratio of k_{cat}/K_m, and the fidelity of nucleotide incorporation was gauged by the frequency of misincorporation (f_{inc}), which was calculated using the following equation:[36]

$$f_{inc} = \frac{(k_{cat} / K_m)_{incorrect}}{(k_{cat} / K_m)_{correct}}$$

Results

Construction and Characterization of Plasmids Housing a Site-specifically Inserted N^2-CEdG or N^2-CMdG

Our previous studies revealed that DinB is the major DNA polymerase involved in the accurate and efficient bypass of N^2-CEdG in *E. coli* cells. The goal of the present investigation is to understand how N^2-CEdG and its structurally related N^2-CMdG are replicated in mammalian cells. To this end, we first constructed double-stranded pTGFP-Hha10 and pMTEX4 vectors housing these lesions (Figures 4.2).

Constructing the lesion-bearing plasmid is among the most challenging steps when a double-stranded shuttle vector is employed for *in vivo* replication experiments. However, the use of the pTGFP-Hha10 and pMTEX4 vectors allows for the relatively easy preparation of the lesion-containing plasmid because the ligation only necessitates the insertion of a single-stranded damage-containing ODN into a gapped double-stranded vector. As depicted in Figure 4.2C (*lane 4*), the strategy facilitated the efficient generation of lesion-containing double-stranded vector, and agarose gel electrophoresis afforded convenient purification of the resulting double-stranded vector from the ligation mixture (Figure 4.2C); we could routinely produce the N^2-CMdG- and N^2-CEdG-containing pTGFP-Hha10 vectors at an overall yield of ~30%.

We further confirmed the incorporation of the lesion-containing insert by employing a restriction digestion/post-labeling assay. In this experiment, the aforementioned double-stranded genomes were digested with EcoRI (the unique EcoRI site is shown in Figure 4.2B). The nascent terminal phosphate groups were removed by

using alkaline phosphatase, and the 5′-termini were subsequently rephosphorylated with [γ-^{32}P]ATP. The linearized vector was further cleaved with NheI (unique NheI site shown in Figure 4.2B), which affords a 52-mer lesion-containing ODN if the ligation is successful (Figure 4.2B). Indeed, denaturing PAGE analysis revealed a distinct 52-mer ^{32}P-labeled fragment, and no shorter fragments were detected (Figure 4.2D), supporting the successful incorporation of the 17-mer lesion-containing insert and the 16-mer unmodified ODN into the gapped construct.

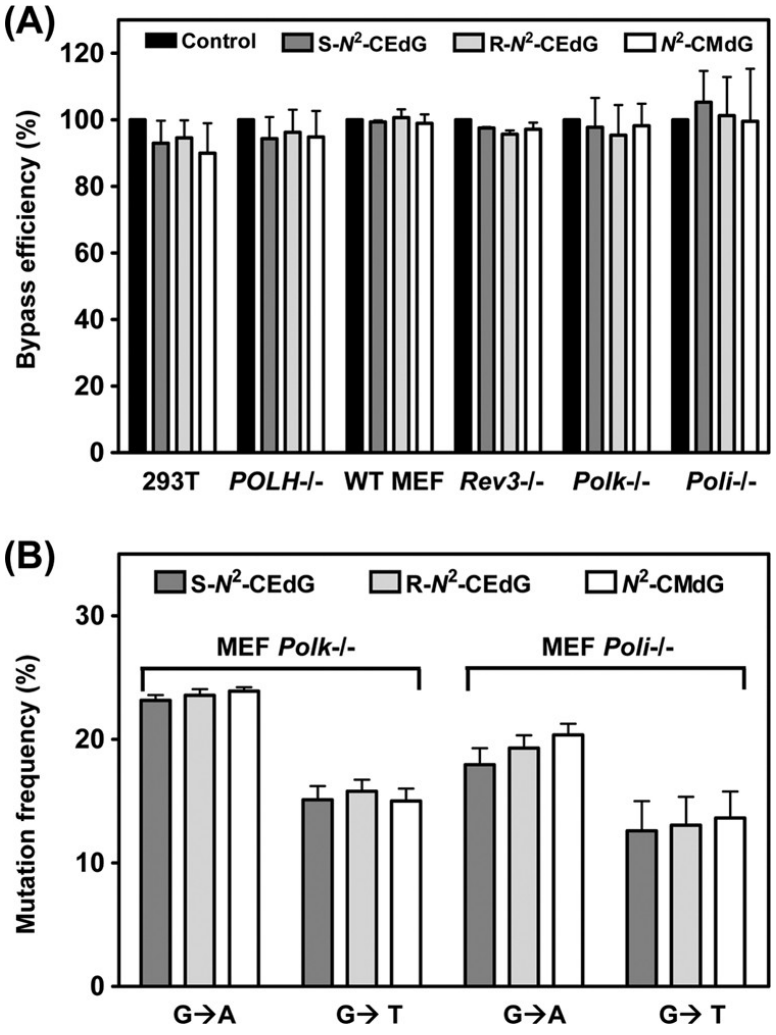
Effect of N^2-CEdG and N^2-CMdG on Efficiency and Fidelity of DNA Replication in Mammalian Cells

When the lesion-containing, double-stranded shuttle vector is replicated in mammalian cells, the undamaged strand may be replicated preferentially over the lesion-carrying strand,[36, 37] rendering it difficult to determine accurately the mutation frequencies. To overcome this difficulty, we employed a similar strategy as reported previously[26, 36, 37] and incorporated an A:A mismatch three nucleotides away from the lesion site (Figures 4.2B and 4.3A). This method facilitated the independent assessment of the products arising from the replication of the lesion-containing strand and its opposing unmodified strand (Figure 4.3A).

The lesion-containing and the control lesion-free vectors were transfected separately into several mammalian cell lines, which include the 293T cells and wild-type MEF cells. To assess the roles of various TLS DNA polymerases in bypassing the above-mentioned N^2-dG lesions, we also evaluated the replication of the lesion-carrying and control lesion-free vectors in immortalized MEF cells deficient in *Polk*,[21] *Rev3*,[22] or *Poli*,[23] as well as C-Tag cells, which are deficient in human *POLH*.[24]

Our results revealed that none of the N^2-dG lesions are mutagenic in human 293T cells or wild-type MEF cells. However, all of them are mutagenic in MEF cells that are deficient in *Polk*. As depicted in Figure 4.3B, replication of A:A mismatch-containing, lesion-free genome (labeled as "control") in wild-type MEF gives two 10-mer products, *i.e.* d(CAAGCTTTGC) (10-mer-AC) and d(CTAGCTTTGC) (10-mer-TC), from the respective replication of the top and bottom strands of the plasmid. The observation of similar amounts of products originating from the replication of the top and bottom strands indicates the lack of repair of the A:A mismatch in the plasmid. The replication of the N^2-CEdG- and N^2-CMdG-harboring genomes in *Polk*-deficient background results in the appearance of two new bands in the gel (Figure 4.3B), which are attributed to d(CTAGTTTTGC) (10-mer-TT) and d(CTAGATTTGC) (10-mer-TA) based on LC-MS/MS monitoring of the complementary 14-mer fragments from the same restriction digestion mixtures (Figures C.3 and 4.3A). In this context, the latter analysis revealed the formation of only the two mutated products as labeled to the right of Figure 3.3B; other mutagenic products are below the detection limit of the LC-MS/MS method. Quantification revealed that the replication of (*S*)-N^2-CEdG-, (*R*)-N^2-CEdG-, and N^2-CMdG-containing genomes in *Polk*-deficient background can give rise to G→A transition at frequencies of 23.3, 23.6, and 23.9%, respectively, and G→T transversion at frequencies of 15.1, 15.8, and 15.0%, respectively (Figure 4.4B). The bypass efficiencies for the three lesions are similar in 293T cells and in wild-type MEF cells, and the deficiency in *Polk* did not compromise the bypass efficiencies. In this regard, the bypass efficiencies for (*S*)-N^2-CEdG, (*R*)-N^2-CEdG, and N^2-CMdG in wild-type MEF are 99.4,

Figure 4.4. Bypass efficiencies (A) and mutation frequencies (B) of dG, (*S*)-N^2-CEdG, (*R*)-N^2-CEdG, and N^2-CMdG lesions in wild-type, *POLH*-, *Rev3*-, *Polk*-, and *Poli*-deficient mammalian cells. *Black*, *dark gray*, *light gray*, and *white columns* represent the results for substrates carrying dG, (*S*)-N^2-CEdG, (*R*)-N^2-CEdG and N^2-CMdG, respectively. The data represent the means and S.D. of results from three independent experiments.



100.6, and 99.0%, respectively, whereas the corresponding bypass efficiencies in the *Polk*-deficient MEF are 97.8, 95.4, and 98.3% (Figure 4.4A).

The above results support that the polymerase κ-mediated bypass of N^2-substituted dG lesions in mammalian cells is accurate. This result is in keeping with our recently published results showing that polymerase IV, the *E. coli* ortholog of mammalian polymerase κ, is the major polymerase involved in the accurate bypass of N^2-CEdG in *E. coli* cells and that the human polymerase κ-mediated *in vitro* nucleotide incorporation opposite N^2-CEdG is both accurate and efficient.[14]

We next assessed whether other TLS DNA polymerase(s) are also involved in the bypass of N^2-CEdG and N^2-CMdG in mammalian cells. Our results revealed that *Poli* deficiency also led to elevated frequencies of G→A (18.0, 19.3, and 20.4% for (*S*)-N^2-CEdG, (*R*)-N^2-CEdG, and N^2-CMdG, respectively) and G→T mutations (12.6, 13.1, and 13.6%, respectively, (Figure 4.4B). In addition, the bypass efficiencies for the three lesions in *Poli*-deficient cells were 105.3, 101.2, and 99.5%, respectively (Figure 4.4A), which were similar to those found in the wild-type MEF cells.

The observation of substantial frequencies of G→A and G→T mutations in *Poli*-deficient MEF suggests that the polymerase ι-mediated nucleotide insertion opposite the above N^2-dG lesions is also accurate. To examine whether this is the case, we performed steady-state kinetic measurements for human polymerase ι-induced nucleotide incorporation opposite the two diastereomers of N^2-CEdG as well as an unmodified dG. The results indeed demonstrated that human polymerase ι preferentially inserts the correct dCMP across the N^2-CEdG lesions (Table 4.1). Importantly, the efficiencies for

Table 4.1. Fidelity of nucleotide incorporation by human polymerases η and ι on N^2-CEdG-containing substrates and the undamaged substrate as determined by steady-state kinetic measurements. k_{cat} and K_m are average values based on three independent measurements.

dNTP	k_{cat}	K_m	k_{cat}/K_m	f_{inc}
	min^{-1}	nM	$nM^{-1}min^{-1}$	
Human polymerase η				
(*S*)-N^2-CEdG-containing substrate				
dATP	$(5.06 \pm 0.01) \times 10^{-2}$	$(3.53 \pm 0.29) \times 10^{3}$	1.43×10^{-5}	0.63
dGTP	$(5.73 \pm 0.37) \times 10^{-2}$	$(2.05 \pm 0.42) \times 10^{4}$	2.79×10^{-6}	0.12
dCTP	$(5.73 \pm 0.37) \times 10^{-2}$	$(2.05 \pm 0.42) \times 10^{4}$	2.79×10^{-6}	0.12
dTTP	$(5.73 \pm 0.37) \times 10^{-2}$	$(2.05 \pm 0.42) \times 10^{4}$	2.79×10^{-6}	0.12
(*R*)-N^2-CEdG-containing substrate				
dATP	$(3.53 \pm 0.01) \times 10^{-2}$	$(2.85 \pm 0.25) \times 10^{3}$	1.24×10^{-5}	0.52
dGTP	$(4.86 \pm 0.01) \times 10^{-2}$	$(5.60 \pm 0.62) \times 10^{4}$	8.69×10^{-7}	0.069
dCTP	$(2.61 \pm 0.02) \times 10^{-2}$	$(1.10 \pm 0.09) \times 10^{3}$	2.37×10^{-5}	1
dTTP	$(1.57 \pm 0.02) \times 10^{-2}$	$(5.62 \pm 0.36) \times 10^{2}$	2.79×10^{-5}	1.2
Control dG-containing substrate				
dATP	$(4.33 \pm 0.01) \times 10^{-2}$	$(9.27 \pm 0.52) \times 10^{3}$	4.67×10^{-6}	0.00013
dGTP	$(1.53 \pm 0.003) \times 10^{-2}$	$(2.52 \pm 0.44) \times 10^{3}$	6.07×10^{-6}	0.00017
dCTP	$(3.57 \pm 0.04) \times 10^{-2}$	$(9.92 \pm 0.72) \times 10^{-1}$	3.61×10^{-2}	1
dTTP	$(3.22 \pm 0.001) \times 10^{-2}$	$(2.56 \pm 0.21) \times 10^{4}$	1.26×10^{-6}	3.5×10^{-5}
Human polymerase ι				
(*S*)-N^2-CEdG-containing substrate				
dCTP	$(1.16 \pm 0.01) \times 10^{-2}$	$(2.23 \pm 0.38) \times 10^{4}$	5.24×10^{-7}	1
(*R*)-N^2-CEdG-containing substrate				
dCTP	$(1.07 \pm 0.02) \times 10^{-2}$	$(1.97 \pm 0.12) \times 10^{4}$	5.43×10^{-7}	1
Control dG-containing substrate				
dCTP	$(1.32 \pm 0.04) \times 10^{-2}$	$(1.06 \pm 0.08) \times 10^{4}$	1.24×10^{-6}	1

dCMP incorporation opposite the two N^2-CEdG derivatives are only slightly lower than that for dCMP insertion opposite an unmodified dG (Table 4.1). This is consistent with the previous observations showing that polymerase ι-mediated nucleotide insertion opposite other minor groove N^2-dG adducts is both accurate and efficient.[38, 39]

We next assessed the role of polymerases η and ζ on the bypass of N^2-CEdG and N^2-CMdG in mammalian cells. It turned out that the replication of N^2-CEdG and N^2-CMdG-bearing plasmids in *POLH*- or *Rev3*-deficient cells did not give rise to detectable mutations. These data underscored the lack of involvement of polymerase η or ζ in bypassing those lesions in a background that is proficient in *Polk* or *Poli*.

Discussion

We developed a shuttle vector-based method for assessing quantitatively how (*S*)-N^2-CEdG, (*R*)-N^2-CEdG, and N^2-CMdG perturb the efficiency and fidelity of DNA replication in mammalian cells. With the use of this method and mammalian cells that are deficient in TLS DNA polymerases, we investigated the roles of various mammalian TLS DNA polymerases in bypassing these lesions. The most important finding made from the present study is that, in mammalian cells, the polymerase κ- and polymerase ι-mediated nucleotide incorporation opposite N^2-CEdG and N^2-CMdG is error-free. This result reinforced the notion that N^2-substituted dG lesions are cognate substrates for DinB DNA polymerases,[12, 14] and it also ascertained that polymerase ι, a Y-family polymerase present only in higher eukaryotes,[4] accurately bypasses these lesions.

Our current study also underscored that, in a genetic background lacking polymerase κ or ι, some other polymerase(s) can also bypass the above two lesions, though with compromised fidelity. In this context, our mutagenesis data revealed substantial frequencies of dAMP and dTMP misincorporation opposite the N^2-dG lesions when the replication was performed in *Polk*- or *Poli*-deficient cells. The *in vitro* steady-state kinetics experiments with purified human polymerase η illustrated that, aside from inserting the correct nucleotide (dCMP), the polymerase can incorporate dAMP and dTMP at efficiencies that are similar or slightly lower than for dCMP insertion (Table 4.1), suggesting that polymerase η might be responsible for the mutagenic bypass of the N^2-dG lesions in a *Polk*- or *Poli*-deficient background. This finding is reminiscent of what was recently observed for the bypass of *cis*,*syn*-cyclobutane thymine dimers in mammalian cells. In this regard, polymerase η can perform the accurate and efficient nucleotide insertion opposite the lesion;[40] in the absence of polymerase η, the lesion also can be bypassed by other TLS polymerases, but with poor accuracy.[40, 41] In addition, a previous report showed that another N^2-modified 2′-deoxyguanosine derivatives, N^2-benzo[α]pyrene-guanine, can be bypassed efficiently and accurately by polymerase κ in wild-type MEF cells; however, the bypass of this adduct was error-prone in the absence of polymerase κ.[42] Along this line, polymerase κ was found to be important to protect mammalian cells against the lethal and mutagenic effects of benzo[α]pyrene.[43]

To further examine whether polymerase η was involved in mutagenic bypass of the N^2-dG lesions in cells, we employed siRNA to knock down the *POLK* gene in *POLH*-deficient cells to determine whether the genetic suppression of *POLK* in *POLH*-deficient

background could abrogate the G→A and G→T mutations found in *Polk*$^{-/-}$ MEFs. As a control experiment, we also treated 293T cells with *POLK* siRNA while reasoning that siRNA-induced depletion of polymerase κ in 293T cells should recapitulate the findings made for *Polk*-deficient MEFs. However, no mutagenic products could be found from replication of the N^2-CEdG- or N^2-CMdG-containing plasmids in 293T cells treated with *POLK* siRNA. This result suggests that, although siRNA treatment results in an ~85% decrease in *POLK* expression as demonstrated by real-time quantitative RT-PCR and Western analysis (Figure C.4), the remaining polymerase κ might be sufficient for bypassing the N^2-carboxyalkyl-dG lesion in transfected plasmids. We also attempted but failed to effectively knock down the expression of *Polh* in the *Pol*κ-deficient MEF cells using siRNA.

Based on the above findings, we put forward a plausible model for the bypass of the N^2-dG lesions in mammalian cells. In wild-type cells, polymerase κ and/or ι insert the correct dCMP opposite the lesion and the resulting N^2-dG:dC base pair at the primer/template junction can be extended by polymerase ζ, a general mismatch extender,[44] or by another TLS polymerase. In cells deficient in *Polk* or *Poli*, another polymerase (*e.g.* polymerase η) can insert a nucleotide opposite the N^2-dG lesion, but the nucleotide incorporation is error-prone, with substantial frequencies of dAMP and dTMP misincorporation. The nascent base pair at the primer/template junction can again be extended by polymerase ζ or another TLS polymerase.

N^2-dG adducts can arise from byproducts of an array of cellular processes, including glycolysis [18, 30] and lipid peroxidation.[45] In addition, there is supporting evidence that

these minor groove adducts may be resistant toward excision repair.[46] In this vein, the N^2-dG adduct is the least abundant dG lesion produced from acetylaminofluorine, but it persists in tissues of animals treated with this carcinogen.[47] The ubiquitous formation of N^2-dG adducts from byproducts of endogenous metabolism and the poor repair of these lesions may justify the need of multiple polymerases (polymerase κ and ι) for their accurate bypass in mammalian cells. In this context, little is known about how N^2-CEdG and N^2-CMdG lesions are repaired, and such studies are currently ongoing in our laboratory.

References

(1) Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature*, *362*, 709-715.

(2) Friedberg E. C., W. G. C., Siede W., Wood R. D., Schultz R. A., Ellenberger T. . (2006) *DNA Repair and Mutagenesis*. ASM Press, Washington, D.C.

(3) Lehmann, A. R., Niimi, A., Ogi, T., Brown, S., Sabbioneda, S., Wing, J. F., Kannouche, P. L., and Green, C. M. (2007) Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst.)*, *6*, 891-899.

(4) Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G. C., Wang, Z., and Woodgate, R. (2001) The Y-family of DNA polymerases. *Mol. Cell.*, *8*, 7-8.

(5) Prakash, S., Johnson, R. E., and Prakash, L. (2005) Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function. *Annu. Rev. Biochem.*, *74*, 317-353.

(6) Friedberg, E. C., Wagner, R., and Radman, M. (2002) Specialized DNA polymerases, cellular survival, and the genesis of mutations. *Science*, *296*, 1627-1630.

(7) Johnson, R. E., Prakash, S., and Prakash, L. (1999) Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Pol η. *Science*, *283*, 1001-1004.

(8) Johnson, R. E., Washington, M. T., Prakash, S., and Prakash, L. (2000) Fidelity of human DNA polymerase η. *J. Biol. Chem.*, *275*, 7447-7450.

(9) Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature*, *399*, 700-704.

(10) Cleaver, J. E. (2005) Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat. Rev. Cancer*, *5*, 564-573.

(11) Ohmori, H., Ohashi, E., and Ogi, T. (2004) Mammalian Pol κ: regulation of its expression and lesion substrates. *Adv. Protein Chem.*, *69*, 265-278.

(12) Jarosz, D. F., Godoy, V. G., Delaney, J. C., Essigmann, J. M., and Walker, G. C. (2006) A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature*, *439*, 225-228.

(13) Choi, J. Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L. A., and Guengerich, F. P. (2006) Translesion synthesis across O^6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J. Biol. Chem.*, *281*, 38244-38256.

(14) Yuan, B., Cao, H., Jiang, Y., Hong, H., and Wang, Y. (2008) Efficient and accurate bypass of N^2-(1-carboxyethyl)-2'-deoxyguanosine by DinB DNA polymerase *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.*, *105*, 8679-8684.

(15) Lone, S., Townson, S. A., Uljon, S. N., Johnson, R. E., Brahma, A., Nair, D. T., Prakash, S., Prakash, L., and Aggarwal, A. K. (2007) Human DNA polymerase κ

encircles DNA: implications for mismatch extension and lesion bypass. *Mol. Cell.*, *25*, 601-614.
(16) Frischmann, M., Bidmon, C., Angerer, J., and Pischetsrieder, M. (2005) Identification of DNA adducts of methylglyoxal. *Chem. Res. Toxicol.*, *18*, 1586-1592.
(17) Schneider, M., Thoss, G., Hubner-Parajsz, C., Kientsch-Engel, R., Stahl, P., and Pischetsrieder, M. (2004) Determination of glycated nucleobases in human urine by a new monoclonal antibody specific for N^2-carboxyethyl-2'-deoxyguanosine. *Chem. Res. Toxicol.*, *17*, 1385-1390.
(18) Li, H., Nakamura, S., Miyazaki, S., Morita, T., Suzuki, M., Pischetsrieder, M., and Niwa, T. (2006) N^2-carboxyethyl-2'-deoxyguanosine, a DNA glycation marker, in kidneys and aortas of diabetic and uremic patients. *Kidney Int.*, *69*, 388-392.
(19) Pluskota-Karwatka, D., Pawlowicz, A. J., Tomas, M., and Kronberg, L. (2008) Formation of adducts in the reaction of glyoxal with 2'-deoxyguanosine and with calf thymus DNA. *Bioorg. Chem.*, *36*, 57-64.
(20) Wang, H., Cao, H., and Wang, Y. (2010) Quantification of N^2-carboxymethyl-2'-deoxyguanosine in calf thymus DNA and cultured human kidney epithelial cells by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope dilution method. *Chem. Res. Toxicol.*, *23*, 74-81.
(21) Bi, X., Slater, D. M., Ohmori, H., and Vaziri, C. (2005) DNA polymerase κ is specifically required for recovery from the benzo[α]pyrene-dihydrodiol epoxide (BPDE)-induced S-phase checkpoint. *J Biol Chem*, *280*, 22343-22355.
(22) Wittschieben, J. P., Reshmi, S. C., Gollin, S. M., and Wood, R. D. (2006) Loss of DNA polymerase ζ causes chromosomal instability in mammalian cells. *Cancer Res.*, *66*, 134-142.
(23) McDonald, J. P., Frank, E. G., Plosky, B. S., Rogozin, I. B., Masutani, C., Hanaoka, F., Woodgate, R., and Gearhart, P. J. (2003) 129-derived strains of mice are deficient in DNA polymerase ι and have normal immunoglobulin hypermutation. *J. Exp. Med.*, *198*, 635-643.
(24) Bullock, S. K., Kaufmann, W. K., and Cordeiro-Stone, M. (2001) Enhanced S phase delay and inhibition of replication of an undamaged shuttle vector in UVC-irradiated xeroderma pigmentosum variant. *Carcinogenesis*, *22*, 233-241.
(25) Cao, H., Jiang, Y., and Wang, Y. (2007) Stereospecific synthesis and characterization of oligodeoxyribonucleotides containing an N^2-(1-carboxyethyl)-2'-deoxyguanosine. *J. Am. Chem. Soc.*, *129*, 12123-12130.
(26) Yuan, B., O'Connor, T. R., and Wang, Y. (2010) 6-Thioguanine and S^6-methylthioguanine are mutagenic in human cells. *ACS Chem. Biol.*, *5*, 1021-1027.
(27) Baker, D. J., Wuenschell, G., Xia, L., Termini, J., Bates, S. E., Riggs, A. D., and O'Connor, T. R. (2007) Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *J Biol Chem*, *282*, 22592-22604.
(28) Hirt, B. (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.*, *26*, 365-369.

(29) Hong, H., Cao, H., and Wang, Y. (2007) Formation and genotoxicity of a guanine-cytosine intrastrand cross-link lesion *in vivo*. *Nucleic Acids Res.*, *35*, 7118-7127.
(30) Yuan, B., and Wang, Y. (2008) Mutagenic and cytotoxic properties of 6-thioguanine, S^6-methylthioguanine, and guanine-S^6-sulfonic acid. *J Biol Chem*, *283*, 23665-23670.
(31) Delaney, J. C., and Essigmann, J. M. (1999) Context-dependent mutagenesis by DNA lesions. *Chem. Biol.*, *6*, 743-753.
(32) Delaney, J. C., and Essigmann, J. M. (2006) Assays for determining lesion bypass efficiency and mutagenicity of site-specific DNA lesions *in vivo*. *Methods Enzymol.*, *408*, 1-15.
(33) Delaney, J. C., and Essigmann, J. M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB *Escherichia coli*. *Proc. Natl. Acad. Sc.i U. S. A.*, *101*, 14051-14056.
(34) Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta C^T)$ Method. *Methods*, *25*, 402-408.
(35) Creighton, S., Bloom, L. B., and Goodman, M. F. (1995) Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol.*, *262*, 232-256.
(36) Levine, R. L., Yang, I. Y., Hossain, M., Pandya, G. A., Grollman, A. P., and Moriya, M. (2000) Mutagenesis induced by a single 1,N^6-ethenodeoxyadenosine adduct in human cells. *Cancer Res.*, *60*, 4098-4104.
(37) Yang, I. Y., Johnson, F., Grollman, A. P., and Moriya, M. (2002) Genotoxic mechanism for the major acrolein-derived deoxyguanosine adduct in human cells. *Chem. Res. Toxicol.*, *15*, 160-164.
(38) Wolfle, W. T., Johnson, R. E., Minko, I. G., Lloyd, R. S., Prakash, S., and Prakash, L. (2006) Replication past a trans-4-hydroxynonenal minor-groove adduct by the sequential action of human DNA polymerases ι and κ. *Mol. Cell. Biol.*, *26*, 381-386.
(39) Pence, M. G., Blans, P., Zink, C. N., Hollis, T., Fishbein, J. C., and Perrino, F. W. (2009) Lesion bypass of N^2-ethylguanine by human DNA polymerase ι. *J Biol Chem*, *284*, 1732-1740.
(40) Yoon, J. H., Prakash, L., and Prakash, S. (2009) Highly error-free role of DNA polymerase η in the replicative bypass of UV-induced pyrimidine dimers in mouse and human cells. *Proc. Natl. Acad. Sci. U. S. A.*, *106*, 18219-18224.
(41) Ziv, O., Geacintov, N., Nakajima, S., Yasui, A., and Livneh, Z. (2009) DNA polymerase ζ cooperates with polymerases κ and ι in translesion DNA synthesis across pyrimidine photodimers in cells from XPV patients. *Proc. Natl. Acad. Sci. U. S. A.*, *106*, 11552-11557.
(42) Avkin, S., Goldsmith, M., Velasco-Miguel, S., Geacintov, N., Friedberg, E. C., and Livneh, Z. (2004) Quantitative analysis of translesion DNA synthesis across a

benzo[α]pyrene-guanine adduct in mammalian cells: the role of DNA polymerase κ. *J Biol Chem*, *279*, 53298-53305.
(43) Ogi, T., Shinkai, Y., Tanaka, K., and Ohmori, H. (2002) Pol κ protects mammalian cells against the lethal and mutagenic effects of benzo[α]pyrene. *Proc. Natl. Acad. Sci. U. S. A.*, *99*, 15548-15553.
(44) Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S., and Prakash, L. (2000) Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions. *Nature*, *406*, 1015-1019.
(45) Marnett, L. J. (2000) Oxyradicals and DNA damage. *Carcinogenesis*, *21*, 361-370.
(46) Zewail-Foote, M., Li, V. S., Kohn, H., Bearss, D., Guzman, M., and Hurley, L. H. (2001) The inefficiency of incisions of ecteinascidin 743-DNA adducts by the UvrABC nuclease and the unique structural feature of the DNA adducts can be used to explain the repair-dependent toxicities of this antitumor agent. *Chem. Biol.*, *8*, 1033-1049.
(47) Yasui, M., Dong, H., Bonala, R. R., Suzuki, N., Ohmori, H., Hanaoka, F., Johnson, F., Grollman, A. P., and Shibutani, S. (2004) Mutagenic properties of 3-(deoxyguanosin-N^2-yl)-2-acetylaminofluorene, a persistent acetylaminofluorene-derived DNA adduct in mammalian cells. *Biochemistry*, *43*, 15005-15013.

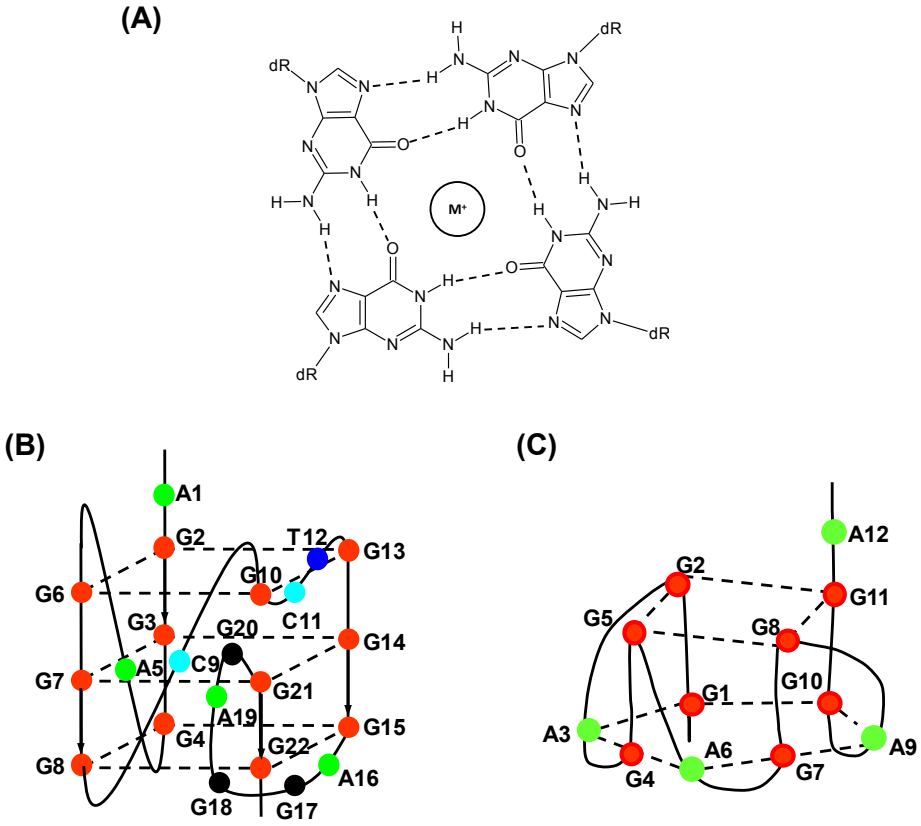
Chapter 5: Replication Across Endogenously Derived G-quadruplex Sequences in *Escherichia coli* Cells

Introduction

During replication and transcription, normally duplex B-form DNA is separated resulting in transiently single-stranded DNA. Under certain conditions, some single-stranded DNA sequences are capable of adopting non-B structures, including triplexes, cruciforms, slipped structures, Z-DNA and G-quadruplexes (G4).[1] These non-B DNA conformations can be mutagenic, causing gene rearrangements and genome instability, and have been implicated in neurological and psychiatric diseases.[2, 3] Amongst the different non-B DNA conformations, G4 structures have received much attention due to their location throughout the genome and their unique physical properties.[4-6]

In vitro biochemical experiments have demonstrated that single-stranded guanine-rich DNA and RNA sequences are able to fold into G4 structures.[7-10] These structures consist of two or more stacked guanine tetrads, i.e., four guanine bases stabilized through Hoogsteen hydrogen-bonding (Figure 5.1A). These tetrads are linked with single-stranded oligonucleotide loops of varying lengths and are further stabilized by electrostatic interactions with monovalent cations located in the central cavity of the tetrad (Figure 5.1A). Depending upon the orientation of the DNA/RNA strands, the G4 structures can adopt a variety of conformations, namely, parallel, antiparallel or some combination of the two.[11] Interestingly, compared to double-stranded DNA, G4 structures are extremely stable (with melting temperatures being 20-30°C higher), though the stability of the G4 structure depends on a variety of factors including sequence length

Figure 5.1. G-quadruplex sequences examined in this study. Structure of G-tetrad stabilized by Hoogsteen hydrogen bonding and a central monovalent cation (A); Intramolecular G4 structure for the *c-Kit* G4 promoter sequence d(AGGGAGGGCGCTGGGAGGAGGG) and the *c-Myb* G4 promoter sequence d(GGAGGAGGAGGA) (B-C).



and composition, size of the loops between the guanines, strand alignment and cations involved in binding.[4, 5, 10]

Computational studies have revealed a preponderance of these guanine-rich nucleotide sequences throughout the genomes of prokaryotes and eukaryotes. For example, there are at least 375,000, 1,400 and 600 potential G4 structures in the nuclear genomes of human,[12] *Saccharomyces cerevisiae*,[13] and *Escherichia coli*, respectively.[14] In addition, these potential G4 sequences are conserved in related species [14-16] and are not randomly located throughout genomes, but instead are concentrated in the telomeres of eukaryotes, in micro- and minisatellite regions, in promoter regions of genes, within ribosomal DNA, near transcription factor binding sites and at mitotic and meiotic double-strand break sites.[13-17] Taken together, this non-random concentration of potential G4 structures in the genome suggests likely roles for these non-canonical structures in gene regulation and chromosome stability.

The promoter region of the *c-Kit* oncogene contains two proximal G4 sequences, designated as c-Kit1 and c-Kit2, which were shown through NMR along with other biophysical techniques to fold into G4 structures under physiological conditions.[18] The two G4 sequences in the *c-Kit* promoter do not share the same sequence, nor is there evidence that they interact with each other. The 22-mer G4 sequence corresponding to c-Kit1 has been well characterized and was shown to fold into a unique, all-parallel stranded G4 structure stabilized by three stacked G-tetrads (Figure 5.1B).[19] Owing to the unique structure adopted by the c-Kit1 G4 sequence, several groups have identified

functional G4 ligands capable of interacting with the c-Kit1 G4 sequence and, upon binding, inhibiting transcription of the downstream gene.[20-22]

The G4 sequence located in the promoter region of the *c-Myb* proto-oncogene consists of three $(GGA)_4$ repeats. The $(GGA)_4$ repeat has previously been well characterized by NMR under physiological conditions,[23, 24] which reveals the folding of $(GGA)_4$ into an intramolecular quadruplex comprised of a G:G:G:G tetrad and a G:(A:)G:(A:) G:(A:)G heptad (T:H) (Figure 5.1C). In addition, two of the three $(GGA)_4$ repeats form a dimer that is stabilized by interactions between the heptads of the two quadruplexes to form a tetrad/heptad/heptad/tetrad (T:H:H:T) quadruplex structure.[23, 24] Nevertheless, i*n vitro* experiments have also shown that one $(GGA)_4$ repeat was adequate to block transcription by RNA polymerase.[24, 25]

Recently several groups have explored the role that G4 sequences and their structures may play in replication. In particular, Lopes et al.[26] recently showed that the G4-prone CEB1 mini-satellite region was capable of forming G4 structures on the leading strand during DNA replication, causing subsequent instability to the genome. Additionally, Bétous et al.[27] revealed the involvement of translesion synthesis (TLS) DNA polymerases in processing G4 sequences from the promoter region of *c-Myc* gene. To further investigate the role of G4 sequences on DNA replication, we inserted one G4 sequence from the *c-Kit* promoter region and one $(GGA)_4$ repeat from the *c-Myb* promoter region into single-stranded M13 genomes and assessed how these two sequences are recognized by DNA replication machinery in *E. coli* cells.

Materials and Methods

Oligodeoxyribonucleotides (ODNs) used in this study were purchased from Integrated DNA Technologies (Coralville, IA). [γ-^{32}P]ATP was obtained from Perkin Elmer Life Sciences (Waltham, MA). Chemicals unless otherwise noted were obtained from Sigma-Aldrich (St. Louis, MO). M13mp7(L2) and wild-type *E. coli* strains were kindly provided by Prof. John M. Essigmann, and polymerase-deficient AB1157 strains [Δ*pol B1*::spec (pol II-deficient), Δ*dinB* (pol IV-deficient), Δ*umuC*::kan (pol V-deficient), Δ*umuC*::kan Δ*dinB*::spec (pol IV-, pol V-double knock-out) and Δ*umuC*::kan Δ*dinB*::Δ*pol B1*::spec (pol IV-, pol V-, pol II-triple knock-out)] were generously provided by Prof. Graham C. Walker.[28] Wild-type RW118 *E. coli* strain and the isogenic helicase-deficient strains [*ΔrecD::kan* (RecD-deficient) and *ΔrecQ::kan* (RecQ-deficient)] were generously provided by Dr. Roger Woodgate (NIH, Rockville, MD).

Construction of G4-containing Single-stranded M13 Genome

Genomes containing G4 sequences and their mutated counterparts were constructed following previously developed procedures.[29] Briefly, 20 pmol of single-stranded M13mp7(L2) genomes were linearized with EcoRI. Two scaffolds spanning each end of the cleaved vector and the ODN sequence were annealed with the linearized vector. The inserts with G4 and control sequences were 5' phosphorylated and ligated into the annealed vectors. The insert sequences and scaffolds used for genome construction are listed in Table D.1. In this context, we mutated two guanine residues in c-Kit-G4 sequence that are essential for G4 structure formation to thymines to give the c-Kit-control sequence. Similarly we mutated four guanine residues required for G4

structure formation to cytosines to yield c-Myb-control sequence. To determine bypass efficiency across the G4 sequences, we also prepared a competitor genome by inserting an unmodified ODN into the linearized M13 genome that was either three bases shorter (for *c-Kit*-G4) or longer (for *c-Myb*-G4) than the G4 and control sequences (Table D.1). Following ligation, the residual linearized single-stranded M13 vector and scaffolds were degraded by addition of T4 DNA polymerase. The reaction mixtures were purified using E.Z.N.A. cycle-pure kit (Omega, Norcross, GA) to remove the enzymes, salts and degraded DNA. The constructed G4 and control genomes were then normalized against the respective competitor genome to accurately define the molar ratio of the G4 and control genomes over the competitor genome, following published procedures.[29]

Transfection of G4-containing Single-stranded M13 Genome.

The G4, or the corresponding control genome was mixed with the competitor genome at a molar ratio of 3/1, transformed into wild-type and the above-described polymerase-deficient AB1157 *E. coli* cells via electroporation. Constructed vectors were also transformed into wild-type RW118 *E. coli* cells and the aforementioned isogenic cells deficient in RecD or RecQ. Following transformation, the plasmids were allowed to replicate in the host cells for 6 hr. In this vein, it is worth nothing that the AB1157 and RW118 strains are proficient in nucleotide excision repair. Following 6 hr of replication, the AB1157 and RW118 cells were pelleted and the supernatant, which contained the progeny phages, was collected. A second amplification was performed by infecting the SCS110 *E. coli* cells with viable progeny phage so as to increase the progeny/parent genome ratios by several orders of magnitude. The infected SCS110 *E. coli* cells were

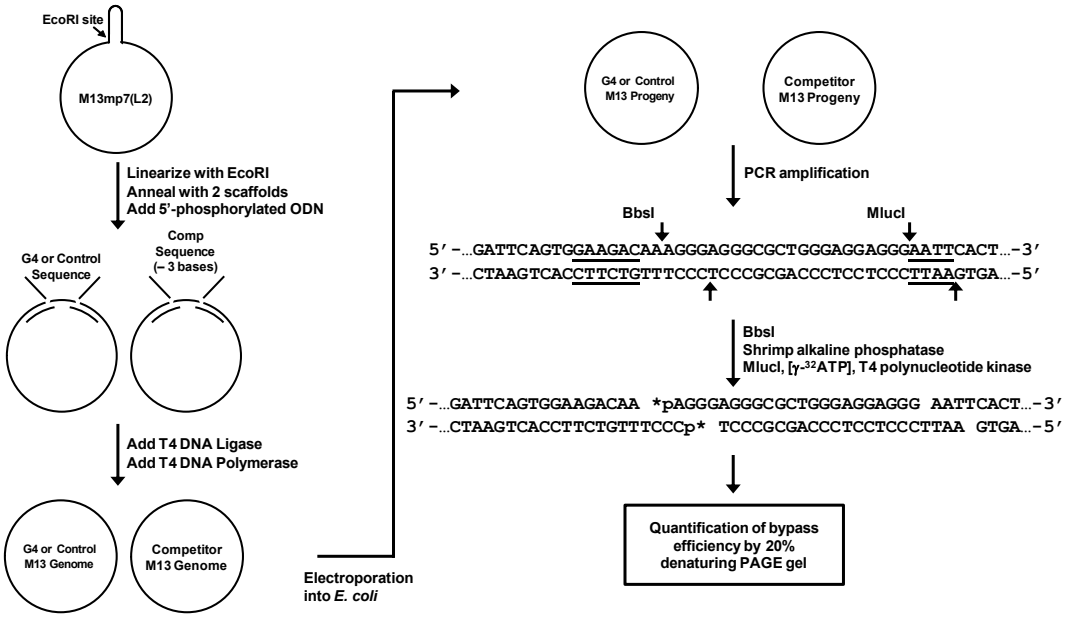
allowed to incubate for 5 hr, after which the single-stranded DNA was extracted from the progeny phage.

Competitive Replication and Adduct Bypass (CRAB) Assay for the Determination of Bypass Efficiency of G4-containing Single-stranded M13 Genomes

CRAB assays were carried out following previously published procedures with some modifications [29]. Briefly, PCR amplification was performed using primers 5′-YGCTATGACCATGATTCAGTGGAAGAC -3′ and 5′-YTTGGGTAACGCCAGGGTTT -3′ (Y is an amino group) which flanked G4, control and corresponding competitor sequences. To assess the bypass efficiency, restriction sites were designed such that the entire or part of the G4 or control sequence could be used a read-out. For the *c-Kit*-G4 genomes, PCR products were treated sequentially with BbsI and shrimp alkaline phosphatase (SAP), followed by ^{32}P-post-labeling and MlucI cleavage. However, for *c-Myb*-G4 genomes, labeling of the G4-containing strand resulted in non-specific cleavages which interfered with quantification. Therefore, the opposite strand was labeled to monitor the bypass of the *c-Myb*-G4 genomes, where the PCR products were treated sequentially with MlucI and SAP, followed by ^{32}P- post-labeling and BbsI cleavage. Reactions were terminated by addition of formamide gel loading buffer containing xylene cyanol FF and bromophenol blue dyes. The mixture was loaded onto a 20% (acrylamide:bisacrylamide = 19:1) denaturing gel, and the resolved ODN fragments quantified by phosphorimaging analysis. For *c-Kit*-G4, restriction digestion of the PCR products from G4 and control progeny genomes products led to radiolabeled 22-mer ODN DNA fragments d(AGGGAGGGCGCTGGGAGGAGGG) (G4) and

d(AGGGAGGGCTCTGTGAGGAGGG) (control), whereas restriction digestion of the competitor PCR product resulted in a 19-mer radiolabeled ODN, d(AGGGAGGGCTTGAGGAGGG) (Scheme 5.1). For *c-Myb*-G4, restriction digestion of PCR products from G4 and control progeny genomes PCR products gives rise 12-mer 5'-^{32}P-labeled ODN DNA fragments, d(AATTTCCTCCTC) (G4) and d(AATTTGCTGCTG) (control); while restriction digestion of the competitor PCR product resulted in a 15-mer 5'-^{32}P-labeled ODN, d(AATTTGCTGCTGCTG) (Scheme D.1). The bypass efficiency was determined using the formula % bypass = (G4 signal/competitor signal)/(control signal/competitor signal).[29]

Scheme 5.1. Methods for the determination of cytotoxicity and mutagenicity of *c-Kit*-G4 sequences in *E. coli* cells.



Results

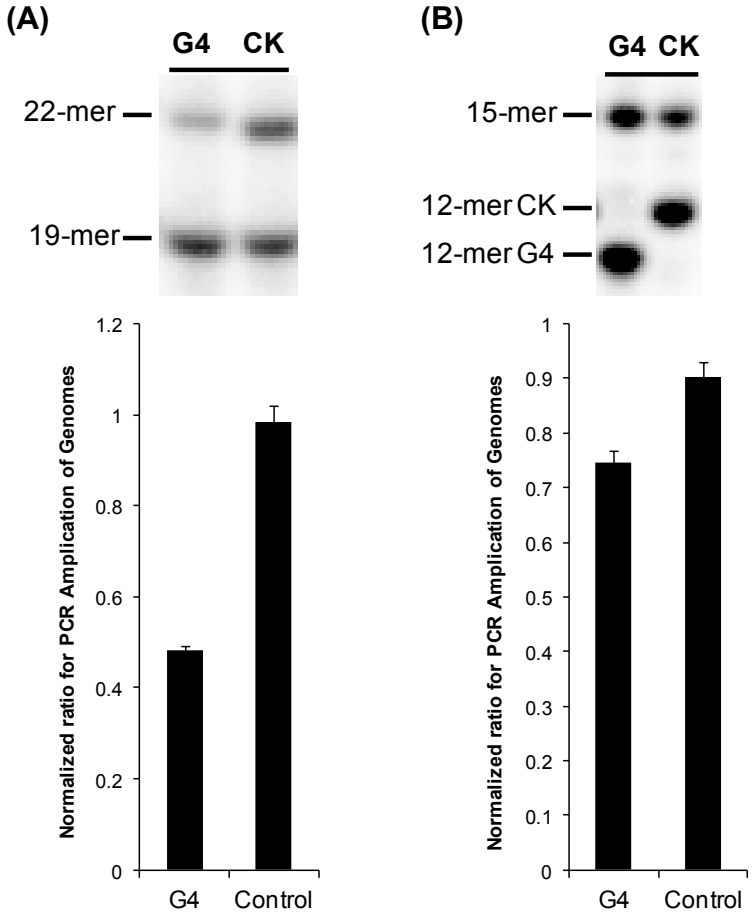
The major objective of the present study was to assess how the presence of G4 structure perturbs DNA replication in *E. coli* cells. Depletion of TLS polymerases was demonstrated to perturb replication of G4 DNA sequence derived from the promoter region of the human *c-Myc* gene [27] and chicken β-globin locus [30], while homologs of the *E. coli* helicases, RecD and RecQ, were shown to play a role in processing G4 sequences from a mouse immunoglobulin locus.[31] Here we systematically investigated the roles of TLS polymerases and DNA helicases in the replicative bypass of two G4 sequences derived from the *c-Kit* and *c-Myb* promoters by transforming single-stranded M13 genomes into *E. coli* cells proficient and deficient in TLS polymerases or helicases. In this respect, both G4 sequences were found to fold into secondary G4 structures *in vitro*.[19, 23, 32]

Replication of c-Kit-G4-containing M13 genome in E. coli cells

One caveat of the CRAB assay is an assumption that both the competitor genome and the G4 or control genomes will be amplified at equal efficiencies during PCR. However, potential G4 structure formation may also alter the efficiency of PCR amplification. In this vein, Gomez et al.[33] developed a PCR stop assay in order to investigate ligand binding to G4 structures *in vitro* because of the ability of G4 structures to impede DNA synthesis during PCR. To account for the possibility of differential amplification of the G4 genome compared to the competitor genome, we performed a second normalization of the genomes. Similar to the initial normalization to derive the true molar ratios of the G4 or control genomes to the corresponding competitor genome,

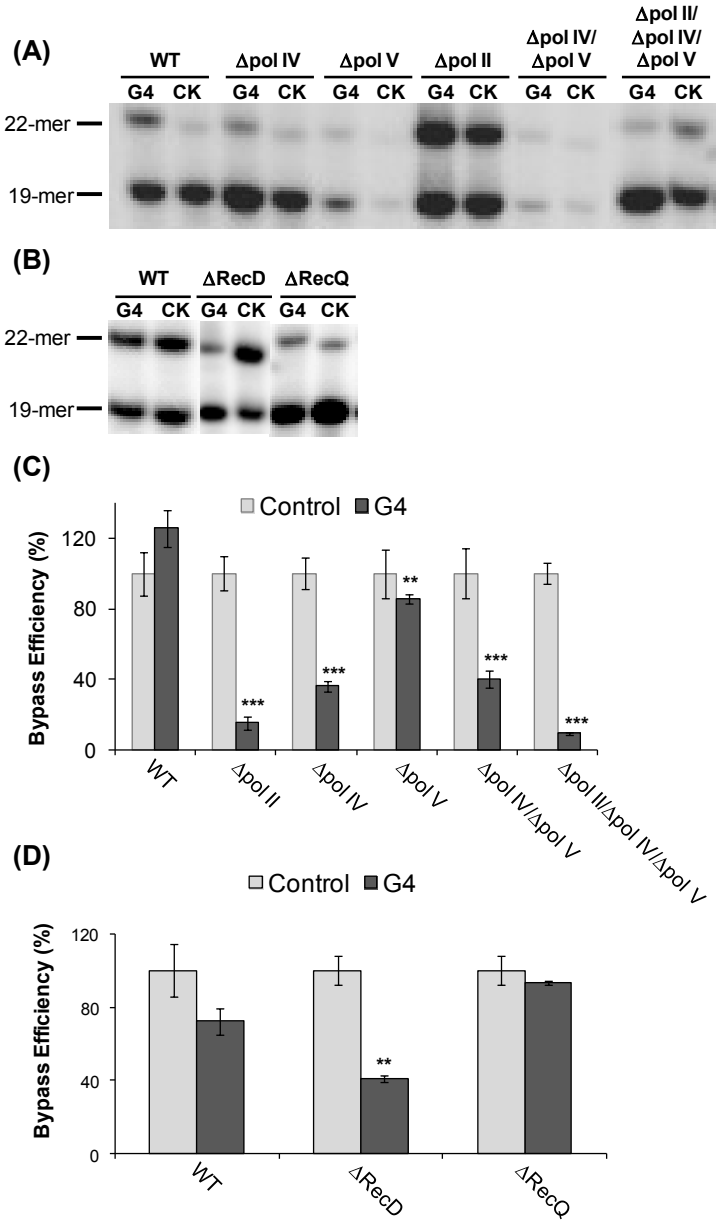
we mixed equal molar amounts of G4 and competitor genomes and subjected them to the same PCR amplification and restriction digestion procedures as those used to interrogate the progeny phage arising from the replication studies. In this way, we can account for any unequal PCR amplification of G4 or competitor genomes. The same procedures were carried out to assess if there was a bias in amplification of the control and competitor genomes. As can be seen in Figure 5.2A, when equal amounts of *c*-Kit-G4 and competitor genomes are PCR amplified, the competitor genome was amplified 2-fold more than the G4 genome, whereas the control and competitor genomes were amplified at nearly equal efficiency.

Figure 5.2. Normalized genome ratios for PCR amplification for the constructed genomes containing *c-Kit-G4* (A) and *c-Myb-G4* (B). In panel (A), the 22-mer bands correspond to the G4 or control sequence while the 19-mer bands correspond to the competitor sequence. In panel (B), the 15-mer bands correspond to the competitor sequence and the 12-mer bands correspond to the G4 and control sequences.



Our results showed that the bypass efficiency for the *c-Kit*-G4 genome was similar to the control genome (Figure 5.3C). Removal of Pol IV and Pol II in the isogenic AB1157 background, however, resulted in significant decreases in bypass efficiencies for *c-Kit*-G4 by ~3.5- and ~8 folds, respectively (Figure5.3C). Further deletion of Pol V in the Pol IV-deficient background did not lead to additional reduction in bypass efficiency. Moreover, we observed an even lower bypass efficiency in the triple-knockout strain (Figure 5.3C). This result suggests a possible role for Pol IV and Pol II in processing the *c-Kit*-G4 sequence during DNA replication in *E. coli* cells.

Figure 5.3. Bypass efficiencies for *c-Kit*-G4 and its control genome in wild-type; pol IV-, pol V- and pol II-deficient; double-knockout, pol IV- and pol V-deficient; and triple-knockout, pol II-, pol IV- and pol V-deficient AB1157 *E. coli* cells (A&C) and in wild-type; RecD- and RecQ-deficient RW118 *E. coli* cells (B&D). In both panels the dark grey and light grey columns represent the results for genomes containing the G4 and control, respectively. p-values in are calculated between the G4 genome bypass efficiencies in wild-type and deficient cells. The 22-mer bands correspond to the G4 and control sequences, while the 19-mer bands correspond to the competitor sequence.



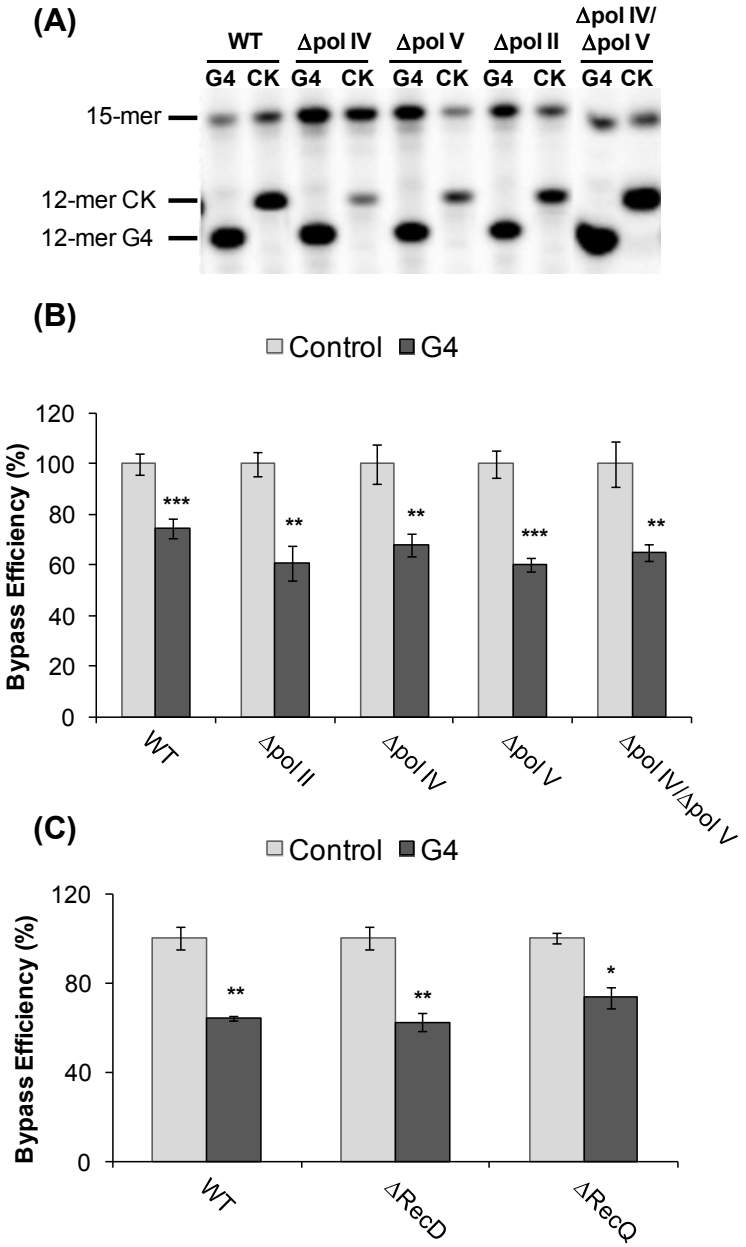
To interrogate the role that DNA helicases play in processing the G4 sequences, we examined how deficiency in RecQ or RecD affects the replicative bypass of the G4 sequences in *E. coli* cells. Both helicases have been implicated in unwinding naturally occurring non-B DNA.[31, 34, 35] Our results revealed a somewhat lower bypass efficiency (~72%) for the *c-Kit*-G4 genome relative to the control genome in wild-type RW118 *E. coli* cells (Figure 5.3D). Removal of the RecQ helicase in the isogenic background resulted in a slight increase in bypass efficiency for the *c-Kit*-G4 genome (by ~30%), whereas the removal of the RecD helicase resulted in a significant reduction in bypass efficiency (by ~1.8 fold, Figure 5.3D). This result suggests that the RecD helicase likely plays a role in processing G4 sequences in *E. coli* cells.

Replication of c-Myb-G4-containing M13 genome in E. coli cells

Similar to the *c-Kit*-G4 genomes, we performed a second normalization to account for any unequal PCR amplification of the *c-Myb*-G4 genome compared to the competitor genome. As can be seen in Figure 5.2B, compared to the *c-Myb*-G4 genome, the competitor genome was amplified ~1.3-fold more efficiently during PCR.

Our CRAB assay results showed that a single $(GGA)_4$ repeat from the *c-Myb* promoter was only slightly blocking to DNA replication in wild-type AB1157 *E. coli* cells, with the bypass efficiency being ~75% (Figure 5.4B). Different from what we observed for the *c-Kit*-G4, removal of pol II, pol IV or pol V in the isogenic AB1157 *E. coli* cells did not appreciably affect the bypass efficiency of the *c-Myb*-G4 sequence (Figure 5.4B). Similarly, in wild-type RW118 *E. coli* cells, DNA replication was only moderately blocked, with the bypass efficiency being ~ 65% (Figure 5.4C). Removal of

Figure 5.4. Bypass efficiencies for *c-Myb*-G4 and its control genome in wild-type; pol II-, pol IV- and pol V-deficient; and double-knockout, pol IV- and pol V-deficient AB1157 *E. coli* cells (A&B) and in wild-type; RecD- and RecQ-deficient RW118 *E. coli* cells (C). In both panels the dark grey and light grey columns represent the results for genomes containing the G4 and control, respectively. p-values were calculated between the G4 and its control genome for each cell line.



the DNA helicase RecD or RecQ did not alter the bypass efficiency (Figure 5.4B), suggesting that the *c-Myb*-G4 sequence modestly blocks DNA replication, but can be resolved without involvement of other polymerases or helicases.

Discussion

Over 40% promoters of all human genes contain at least one G4 sequence, and a large number of these sequences are present in the promoters of oncogenes.[36, 37] Many of these G4 sequences are conserved among different species and have been shown to fold into G4 structures.[36] The proto-oncogenes and oncogenes containing these G4 structures are often involved in cellular functions such as self-sufficiency, cellular growth, apoptosis, angiogenesis, proliferation and replication, and altered expression of these genes is a hallmark of cancer.[36] Additionally, regions within promoters that are considered nuclease-hypersensitive are enriched with G4 sequences (~230-fold higher than the bulk genome).[37] Despite a likely role for G4 structures in the regulation of transcription based upon their locations, not much is known about how these non-B DNA structures affect DNA replication. During DNA replication, duplex DNA becomes transiently single-stranded, allowing for the possible folding of G4 sequences into secondary structures, which may impede progression of the replication fork. Therefore, we interrogated the effects that two G4 sequences located in the promoter regions of two genes have on DNA replication in *E. coli* cells.

Our results revealed that, relative to their respective mutant sequences, G4 sequences derived from the promoters of both *c-Kit* and *c-Myb* genes were capable of

impeding DNA replication. Replication opposite the *c-Myb*-G4 sequence was only slightly attenuated, with an average decrease in bypass efficiency being 20-30% in all cell lines relative to the control sequence incapable of folding into G4 structure. This relatively small reduction in bypass efficiency is likely attributed to the single quadruplex structure, T:H (Figure 5.1C), adopted by the $(GGA)_4$ *c-Myb* sequence. As discussed earlier, in human cells the *c-Myb* promoter contains three consecutive $(GGA)_4$ repeats in which two of the three $(GGA)_4$ repeats stabilize each other to form a T:H:H:T intramolecular quadruplex (Figure 5.1C).[23, 24] While previous studies have shown that one $(GGA)_4$ repeat is adequate to impede transcription, the presence of multiple $(GGA)_4$ repeats results in a substantial decrease in transcription by RNA polymerases.[23, 24]

Bypass efficiencies for the G4 sequence derived from the *c-Kit* promoter were, however, significantly reduced in isogenic AB1157 cells deficient in DNA polymerases Pol II (~19%) or Pol IV (~36%) (Figure 5.3C). While this result differs from that obtained for the *c-Myb* G4 sequence, it is not completely surprising. The promoter region of *c-Kit* contains two G4 sequences capable of folding into secondary structures; however, unlike the *c-Myb* promoter, there is little evidence to suggest that these two G4 sequences stabilize each other. Treatment of HeLa cells, which contain multiple copies of the promoter of the human *c-Myc* gene with only one G4 sequence, with telomestatin (TMS) and depleted of Pol κ and Pol η could result in an increase in the amount of double-strand breaks.[27] Originally identified as a potent telomerase inhibitor owing to its ability to stabilize telomeric G4 structures,[38] TMS has subsequently been shown to also target and stabilize non-telomeric G4 sequences.[39] In addition, Sarkies et al.[30] showed

that the TLS polymerase, Rev1, was required for replication of G4 sequences in leading strand templates. Taken together, our *c-Kit*-G4 results along with other recent observations suggest that TLS polymerases may play a role in processing naturally occurring G4 DNA.

In addition to TLS polymerases, DNA helicases have been implicated in resolving G4 DNA. Removal of the RecD helicase resulted in a substantial drop in bypass efficiency (~2-fold) for *c-Kit*-G4, whereas removal of the RecQ helicase did not alter the bypass efficiency (Figure 5.3D). Our results are in good agreement with a study by Paeschke et al.,[31] who showed that *S. cerevisiae* Pif1, an *E. coli* RecD homolog, along with five other prokaryotic Pif1 helicases, were much more proficient in unwinding G4 sequences than three different RecQ helicases. While they interrogated helicase unwinding of murine immunoglobulin Sγ2b switch region (TP_{G4}) sequence, we show here that *E. coli* RecD is likely involved in unwinding a wide range of G4 sequences including those located in promoter regions.

References

(1) Wang, G., and Vasquez, K. M. (2006) Non-B DNA structure-induced genetic instability. *Mutat. Res.*, *598*, 103-119.

(2) Bacolla, A., and Wells, R. D. (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol. Carcinog.*, *48*, 273-285.

(3) Zhao, J., Bacolla, A., Wang, G., and Vasquez, K. M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell Mol. Life Sci.*, *67*, 43-62.

(4) Bochman, M. L., Paeschke, K., and Zakian, V. A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.*, *13*, 770-780.

(5) Lipps, H. J., and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, *19*, 414-422.

(6) Han, H., and Hurley, L. H. (2000) G-quadruplex DNA: a potential target for anti-cancer drug design. *Trends Pharmacol. Sci.*, *21*, 136-142.

(7) Sen, D., and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, *334*, 364-366.

(8) Gilbert, D. E., and Feigon, J. (1999) Multistranded DNA structures. *Curr. Opin. Struct. Biol.*, *9*, 305-314.

(9) Neidle, S., and Parkinson, G. N. (2003) The structure of telomeric DNA. *Curr. Opin. Struct. Biol.*, *13*, 275-283.

(10) Lane, A. N., Chaires, J. B., Gray, R. D., and Trent, J. O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, *36*, 5482-5515.

(11) Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, *34*, 5402-5415.

(12) Todd, A. K., Johnston, M., and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, *33*, 2901-2907.

(13) Hershman, S. G., Chen, Q., Lee, J. Y., Kozak, M. L., Yue, P., Wang, L. S., and Johnson, F. B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, *36*, 144-156.

(14) Rawal, P., Kummarasetti, V. B., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S. K., and Chowdhury, S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.*, *16*, 644-655.

(15) Capra, J. A., Paeschke, K., Singh, M., and Zakian, V. A. (2010) G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput. Biol.*, *6*, e1000861.

(16) Nakken, S., Rognes, T., and Hovig, E. (2009) The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. *Nucleic Acids Res.*, *37*, 5749-5756.

(17) Huppert, J. L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.*, *37*, 1375-1384.
(18) Kuryavyi, V., Phan, A. T., and Patel, D. J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res.*, *38*, 6757-6773.
(19) Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D. J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *J. Am. Chem. Soc.*, *129*, 4386-4392.
(20) McLuckie, K. I., Waller, Z. A., Sanders, D. A., Alves, D., Rodriguez, R., Dash, J., McKenzie, G. J., Venkitaraman, A. R., and Balasubramanian, S. (2011) G-quadruplex-binding benzo[a]phenoxazines down-regulate *c-KIT* expression in human gastric carcinoma cells. *J. Am. Chem. Soc.*, *133*, 2658-2663.
(21) Wang, X., Zhou, C.X., Yan, J.W., Hou, J.Q., Chen, S.B., Ou, T.M., Gu, L.Q., Huang, Z.H., Tan, J.H. (2013) Synthesis and evaluation of quinazolone derivative as a new class of *c-KIT* G-quadruplex binding ligands. *ACS Med. Chem. Lett.*
(22) Waller, Z. A. E., Sewitz, S. A., Hsu, S. T. D., and Balasubramanian, S. (2009) A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression. *J. Am. Chem. Soc.*, *131*, 12628-12633.
(23) Matsugami, A., Ouhashi, K., Kanagawa, M., Liu, H., Kanagawa, S., Uesugi, S., and Katahira, M. (2001) An intramolecular quadruplex of $(GGA)_4$ triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.*, *313*, 255-269.
(24) Palumbo, S. L., Memmott, R. M., Uribe, D. J., Krotova-Khan, Y., Hurley, L. H., and Ebbinghaus, S. W. (2008) A novel G-quadruplex-forming GGA repeat region in the *c-myb* promoter is a critical regulator of promoter activity. *Nucleic Acids Res.*, *36*, 1755-1769.
(25) Broxson, C., Beckett, J., and Tornaletti, S. (2011) Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myb gene. *Biochemistry*, *50*, 4162-4172.
(26) Lopes, J., Piazza, A., Bermejo, R., Kriegsman, B., Colosio, A., Teulade-Fichou, M. P., Foiani, M., and Nicolas, A. (2011) G-quadruplex-induced instability during leading-strand replication. *EMBO J.*, *30*, 4033-4046.
(27) Betous, R., Rey, L., Wang, G., Pillaire, M. J., Puget, N., Selves, J., Biard, D. S., Shin-ya, K., Vasquez, K. M., Cazaux, C., and Hoffmann, J. S. (2009) Role of TLS DNA polymerases η and κ in processing naturally occurring structured DNA in human cells. *Mol. Carcinog.*, *48*, 369-378.
(28) Jarosz, D. F., Beuning, P. J., Cohen, S. E., and Walker, G. C. (2007) Y-family DNA polymerases in *Escherichia coli*. *Trends Microbiol.*, *15*, 70-77.
(29) Delaney, J. C., and Essigmann, J. M. (2006) Assays for determining lesion bypass efficiency and mutagenicity of site-specific DNA lesions *in vivo*. *Methods Enzymol.*, *408*, 1-15.
(30) Sarkies, P., Reams, C., Simpson, L. J., and Sale, J. E. (2010) Epigenetic instability due to defective replication of structured DNA. *Mol. Cell*, *40*, 703-713.

(31) Paeschke, K., Bochman, M. L., Garcia, P. D., Cejka, P., Friedman, K. L., Kowalczykowski, S. C., and Zakian, V. A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, *497*, 458-462.
(32) Rankin, S., Reszka, A. P., Huppert, J., Zloh, M., Parkinson, G. N., Todd, A. K., Ladame, S., Balasubramanian, S., and Neidle, S. (2005) Putative DNA quadruplex formation within the human *c-kit* oncogene. *J. Am. Chem. Soc.*, *127*, 10584-10589.
(33) Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J. L., and Riou, J. F. (2004) Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.*, *32*, 371-379.
(34) Liu, J. Q., Chen, C. Y., Xue, Y., Hao, Y. H., and Tan, Z. (2010) G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates. *J. Am. Chem. Soc.*, *132*, 10521-10527.
(35) Wu, Y., Shin-ya, K., and Brosh, R. M., Jr. (2008) FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell. Biol.*, *28*, 4116-4128.
(36) Eddy, J., and Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.*, *34*, 3887-3896.
(37) Huppert, J. L., and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, *35*, 406-413.
(38) Gomez, D., O'Donohue, M. F., Wenner, T., Douarre, C., Macadre, J., Koebel, P., Giraud-Panis, M. J., Kaplan, H., Kolkes, A., Shin-ya, K., and Riou, J. F. (2006) The G-quadruplex ligand telomestatin inhibits POT1 binding to telomeric sequences *in vitro* and induces GFP-POT1 dissociation from telomeres in human cells. *Cancer Res.*, *66*, 6908-6912.
(39) De Cian, A., Cristofari, G., Reichenbach, P., De Lemos, E., Monchaud, D., Teulade-Fichou, M. P., Shin-Ya, K., Lacroix, L., Lingner, J., and Mergny, J. L. (2007) Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proc. Natl. Acad. Sci. U. S. A.*, *104*, 17347-17352.

Chapter 5: Conclusions and future research

In the study described in the first chapter of this dissertation, *in vitro* replication studies with duplex DNA substrates containing site-specifically incorporated O^2-MdT and O^4-MdT were conducted using three different DNA polymerases, the exonuclease-free Klenow fragment of *Escherichia coli* DNA polymerase I (Kf$^-$), human DNA polymerase κ (pol κ) and *Saccharomyces cerevisiae* DNA polymerase η (pol η). Steady-state kinetic measurements and LC-MS/MS analyses showed that Kf$^-$ and pol η preferentially incorporated the correct nucleotide (dAMP) opposite O^2-MdT, while O^4-MdT primarily directed dGMP misincorporation. Our results underscored the limitation of the steady-state kinetic assay in determining how DNA lesions compromise DNA replication *in vitro*. While steady-state kinetic experiments showed that opposite O^2-MdT and O^4-MdT nucleotide insertion mediated by pol κ was highly promiscuous is highly promiscuous, LC-MS/MS analysis of primer extension products demonstrated that pol κ primarily incorrect dGMP opposite both lesions. The results from this study underscored a role for *O*-methylated thymidine lesions as important sources of mutations.

In Chapter 2, the effects that a larger ethyl group located on the O^2-, *N*3- and O^4-positions of thymidine have on DNA replication were examined. Comprehensive replication studies using purified human DNA polymerases (hPol) η, κ, and ι, yeast DNA polymerase ζ (yPol ζ), and the exonuclease-free Klenow fragment (Kf$^-$) of *Escherichia coli* DNA polymerase I were used to study the regioisomeric O^2-, *N*3- and O^4-ethylthymidine lesions (O^2-, *N*3- and O^4-EtdT). Both hPol η and Kf$^-$ able to generate

full-length replication products opposite all three ethylated thymidines, whereas hPol ι stalled after inserting a single nucleotide opposite the lesions. Lesion bypass mediated by hPol κ and yPol ζ differed amongst the three lesions: hPol κ was able to bypass O^2-EtdT, but had great difficulty in bypassing *N*3-EtdT and O^4-EtdT; yPol ζ was modestly blocked by O^4-EtdT, whereas the polymerase was highly hindered by O^2-EtdT and *N*3-EtdT. The LC-MS/MS analysis of the replication product distribution were in good agreement with our gel results and revealed that dGMP was preferentially inserted opposite O^4-EtdT, while the presence of O^2-EtdT and *N*3-EtdT in the template resulted in misincorporation of dTMP and dGMP. Our results suggested that O^2-EtdT and *N*3-EtdT may also contribute to the AT→TA and AT→GC mutations observed in cells and tissues of animals exposed to ethylating agents.

In Chapter 3 of this dissertation, endogenously produced N^2-(1-carboxyethyl)-2′-deoxyguanosine (N^2-CEdG) and N^2-carboxymethyl-2′-deoxyguanosine (N^2-CMdG) were studied using both shuttle vector technology and *in vitro* steady-state assays. From *in vivo* replication experiments a deficiency in *Polk* or *Poli* in mouse embryonic fibroblast (MEF) cells resulted in higher frequencies of G→T and G→A mutations at N^2-CEdG and N^2-CMdG lesion sites. Biochemical steady-state kinetic measurements revealed insertion of the correct nucleotide, dCMP, opposite N^2-CEdG lesions by human DNA polymerase ι. Replication of N^2-CEdG- and N^2-CMdG-bearing plasmids in *POLH*-deficient human cells resulted in no mutations. Together, with steady-state kinetic measurements showing elevated amounts of dTMP and dAMP by yeast polymerase η, our results revealed that, in mammalian cells, these lesion can be accurately and efficiently bypassed by

polymerase κ or ι and that other translesion synthesis polymerase(s) could incorporate nucleotide(s) inaccurately opposite these lesions.

Finally in Chapter 4, the effect that non-B, G-quadruplex (G4) structures have on DNA replication was investigated. Using single-stranded shuttle vectors containing the two G4 sequences from the oncogene *c-Kit* and the proto-oncogene *c-Myb*, *in vivo* replication studies were performed in TLS-deficient and Helicase-deficient *E. coli* cells. The experiments revealed that the two G4 sequences were capable of blocking DNA replication. Additionally, *E. coli* TLS Pols II and IV, along with *E. coli* DNA helicase RecD, were shown to be involved in resolving the *c-Kit*-G4 sequence.

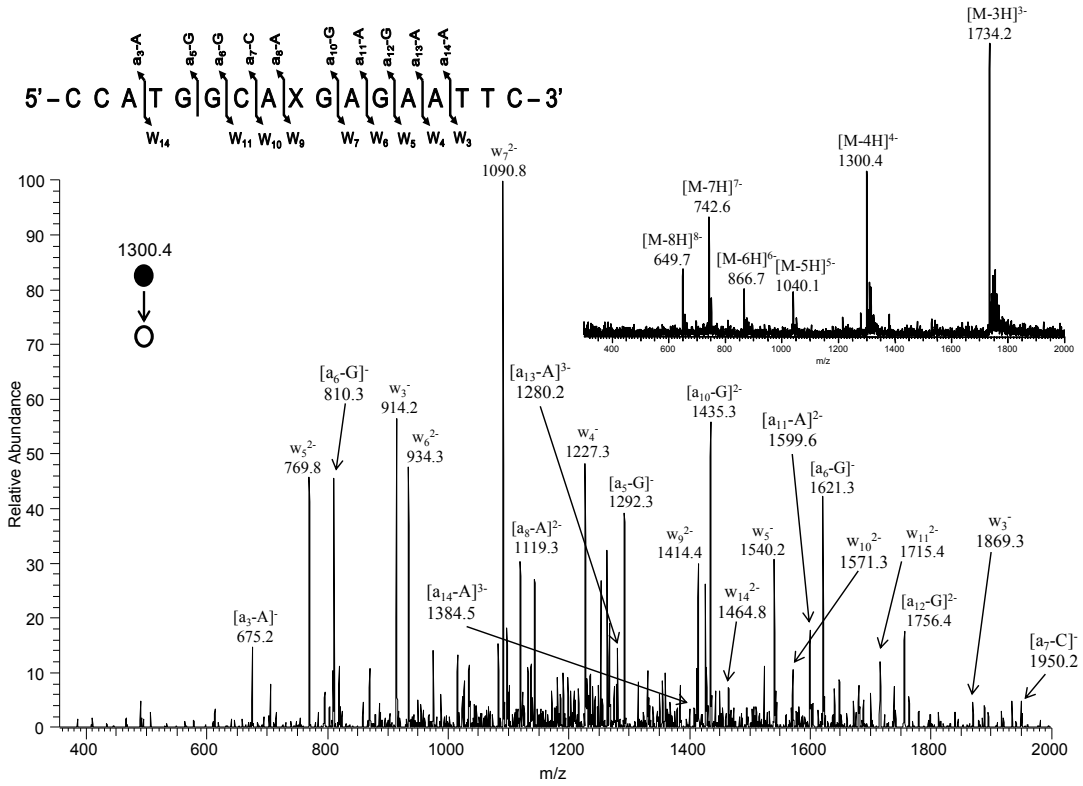
Future research

Future research should further investigate the effect that alkylated thymidines have on DNA replication. Comprehensive *in vitro* studies along with *in vivo* experiments will reveal whether larger alkyl groups located on the thymine nucleobase behave similarly to the smaller methyl and ethyl groups that have already been studied here. Additionally, further studies are needed to explore the role that TLS polymerases and DNA helicases have in processing G4 structures. Examining more G4 sequences will reveal whether the same TLS polymerase is involved in processing all G4 structures or whether the polymerase involved depends upon the particular G4 structure. Other experiments which involve the knock out of both TLS polymerases and DNA helicases simultaneously will also help to reveal whether these TLS polymerases and DNA helicases act together or independently.

Appendix A: Supporting Information for Chapter 2

Figure A.1. Product-ion spectrum of the $[M-4H]^{4-}$ ion (*m/z* 1300.4) of the 17-mer ODN d(CCATGGCA*X*GAGAATTC), where "*X*" is O^2-MdT (A) and O^4-MdT (B). Shown in the inset is the negative-ion ESI-MS for the ODNs.

(A)



(B)

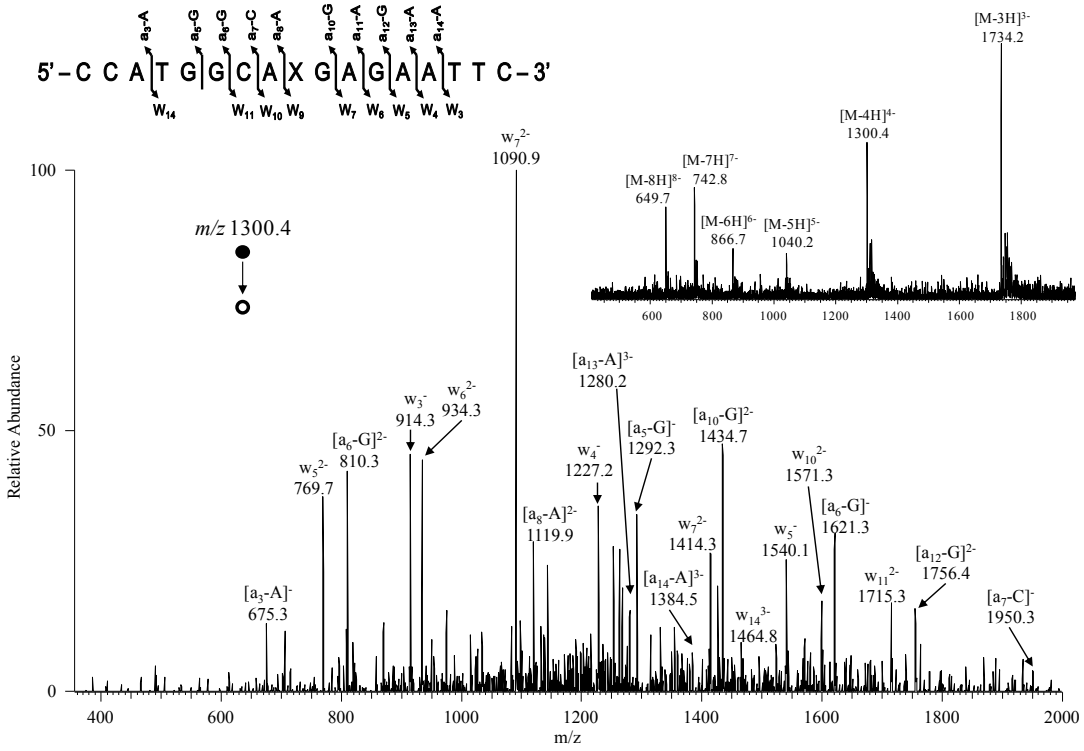
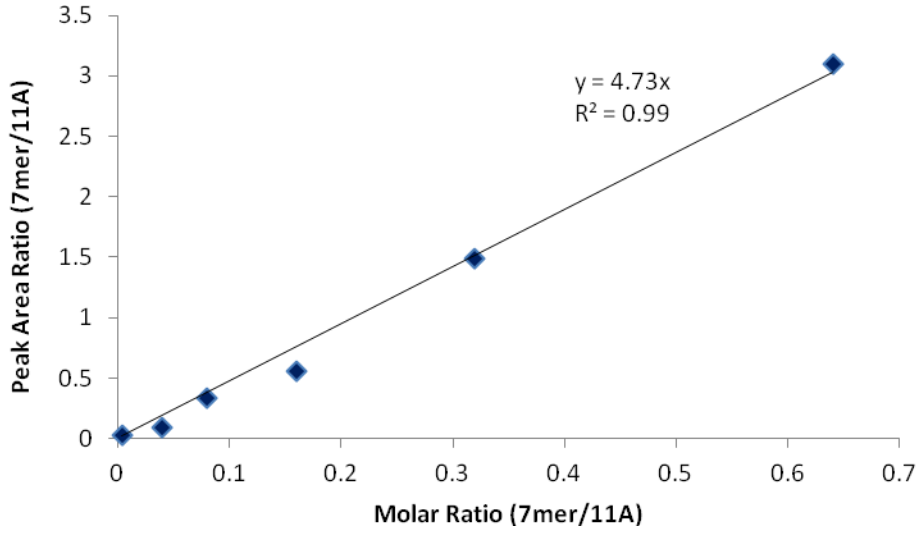
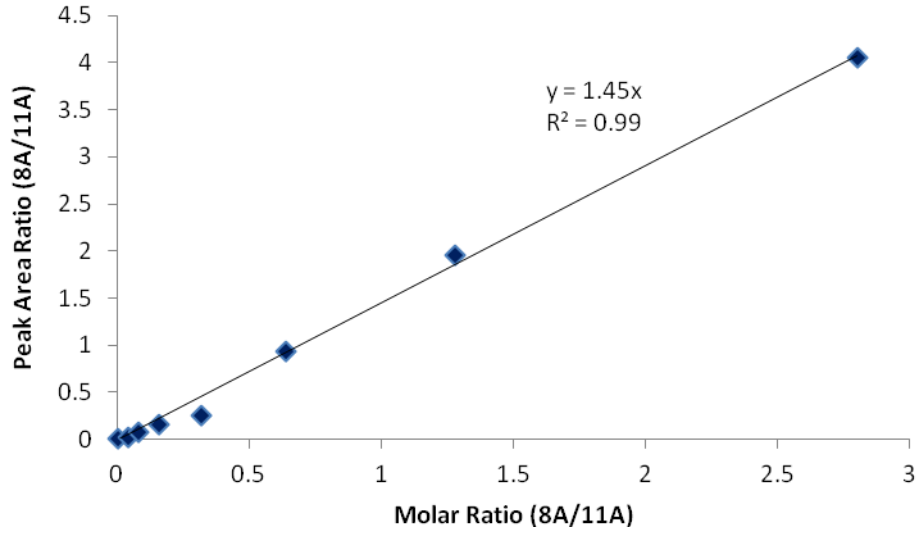

5'-C C A T G G C A X G A G A A T T C-3'
a3-A a6-G a6-G a7-C a6-A a10-G a11-A a12-G a13-A a14-A
W14 W11 W10 W9 W7 W6 W5 W4 W3
m/z 1300.4
w7 2- 1090.9
[M-3H]3- 1734.2
[M-4H]4- 1300.4
[M-7H]7- 742.8
[M-8H]8- 649.7
[M-6H]6- 866.7
[M-5H]5- 1040.2
[a13-A]3- 1280.2
w3- 914.3
w6 2- 934.3
[a6-G]2- 810.3
w5 2- 769.7
[a5-G]- 1292.3
[a10-G]2- 1434.7
w4- 1227.2
w10 2- 1571.3
[a8-A]2- 1119.9
w7 2- 1414.3
[a6-G]- 1621.3
w5- 1540.1
[a12-G]2- 1756.4
[a14-A]3- 1384.5
w11 2- 1715.3
[a3-A]- 675.3
w14 3- 1464.8
[a7-C]- 1950.3
Relative Abundance
m/z

Figure A.2. Calibration curves for the ratios of unextended primer d(AATTCTC) to d(AATTCTCATGC) (i.e., 11A) (A); d(AATTCTCA) (i.e., 8A) to 11A (B); d(AATTCTCG) (i.e., 8G) to 11A (C); d(AATTCTCTGC) (i.e., 10Del) to 11A (D); d(AATTCTCCTGC) (i.e., 11C) to 11A (E); d(AATTCTCTTGC) (i.e., 11T) to 11A (F); d(AATTCTCGTGC) (i.e., 11G) to 11A (G).

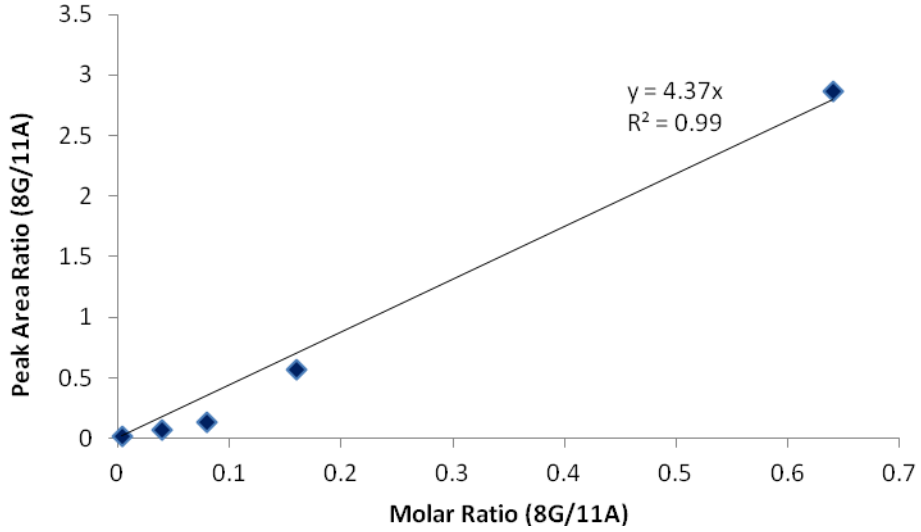
(A)



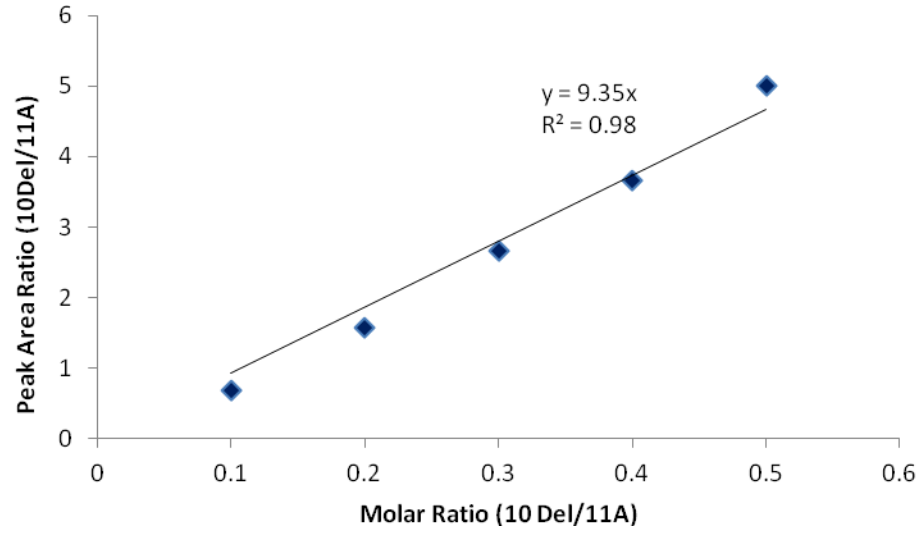
(B)



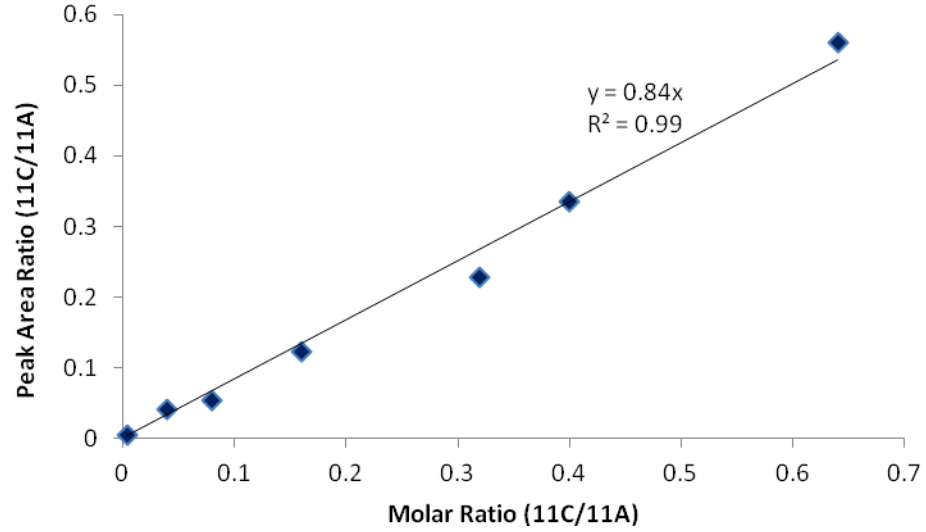
(C)


y = 4.37x
R² = 0.99
Peak Area Ratio (8G/11A)
Molar Ratio (8G/11A)
0
0.5
1
1.5
2
2.5
3
3.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

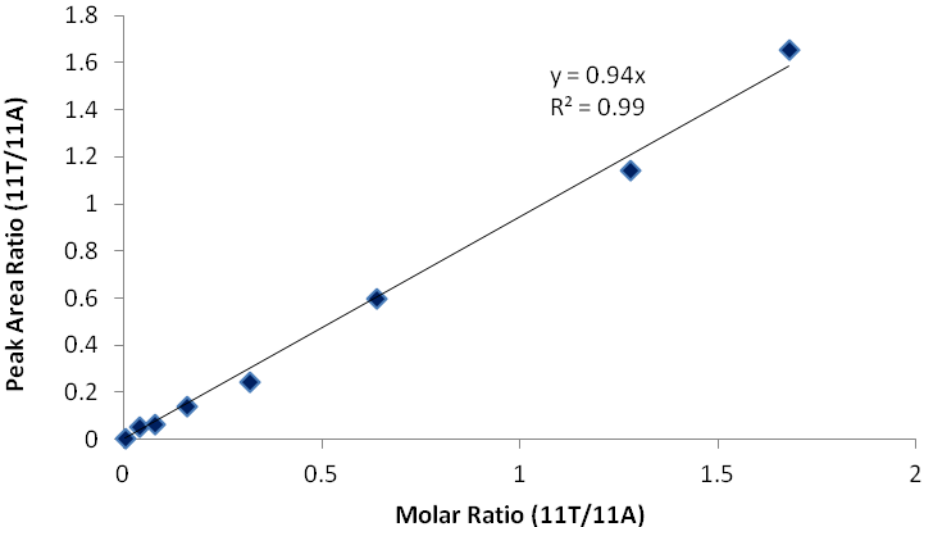
(D)


y = 9.35x
R² = 0.98
Peak Area Ratio (10Del/11A)
Molar Ratio (10 Del/11A)
0
1
2
3
4
5
6
0
0.1
0.2
0.3
0.4
0.5
0.6

(E)


y = 0.84x
R² = 0.99
Peak Area Ratio (11C/11A)
Molar Ratio (11C/11A)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

(F)


y = 0.94x
R² = 0.99
Peak Area Ratio (11T/11A)
Molar Ratio (11T/11A)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
0.5
1.5
2

(G)

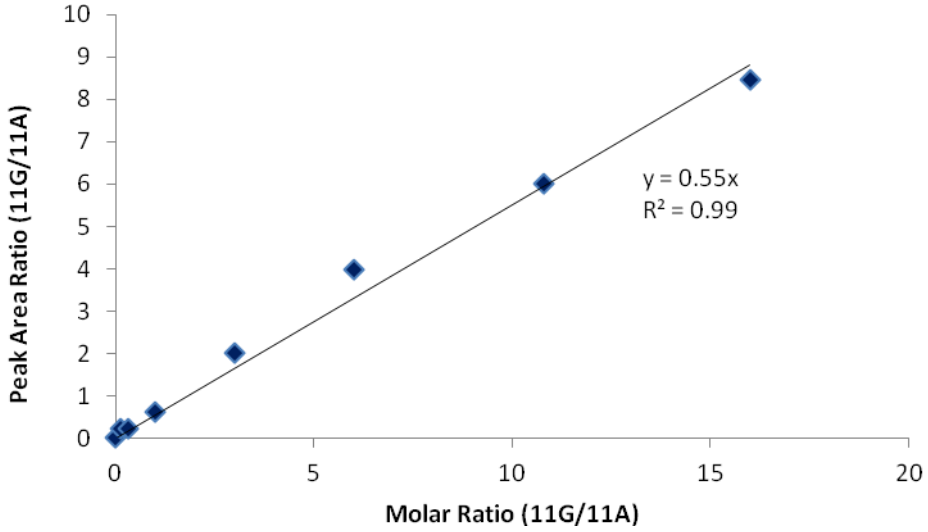

Peak Area Ratio (11G/11A)
Molar Ratio (11G/11A)
y = 0.55x
R² = 0.99
0
1
2
3
4
5
6
7
8
9
10
0
5
10
15
20

Figure A.3. Representative gel images for steady-state kinetic assays measuring nucleotide incorporation opposite undamaged dT (A), O^2-MdT (B), and O^4-MdT (C) with Kf^- (0.7-7 nM). Reactions were conducted in the presence of individual dNTPs with the highest concentrations indicated in the figures. The ratios of dNTP concentrations were 0.5-0.6 between neighboring lanes.

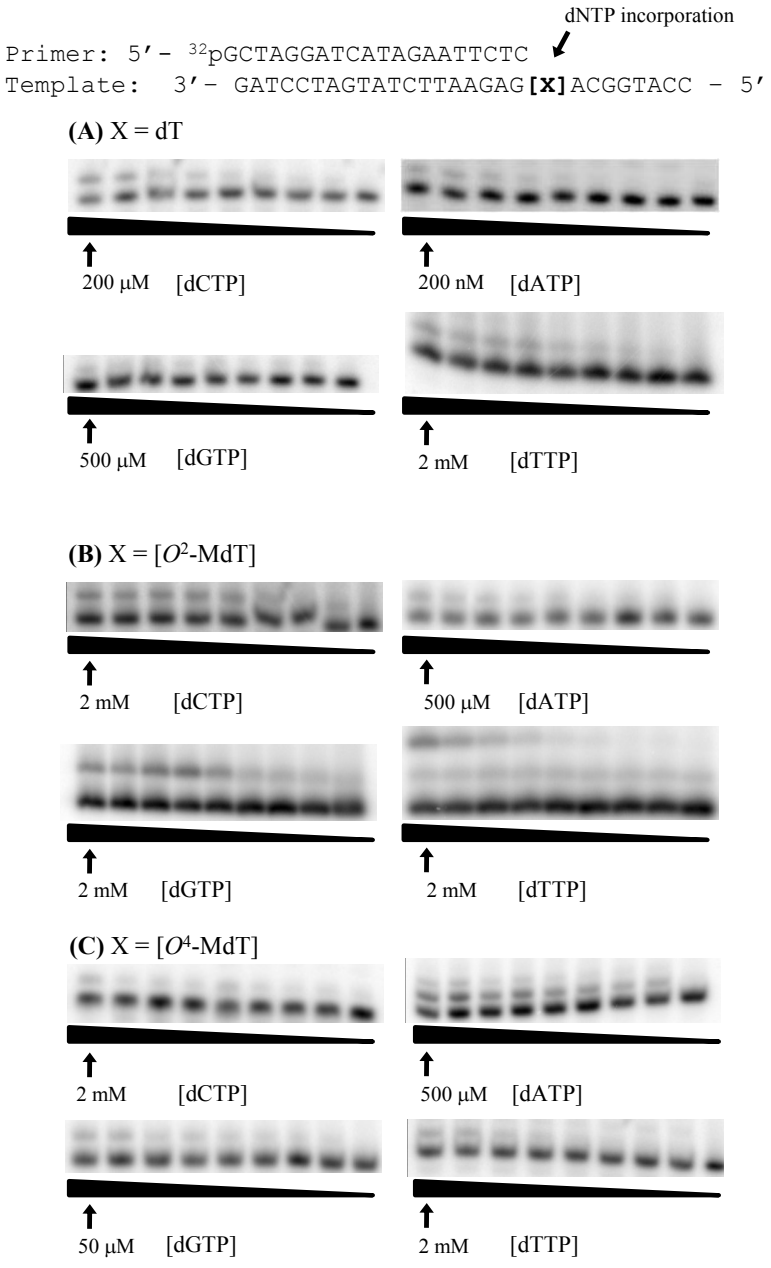


Figure A.4. Representative gel images for steady-state kinetic assays measuring nucleotide incorporation opposite undamaged dT (A), O^2-MdT (B), and O^4-MdT (C) with human DNA polymerase κ (1-5 nM). Reactions were conducted in the presence of individual dNTPs with the highest concentrations indicated in the figures. The concentration ratios of dNTP between neighboring lanes were 0.5-0.6.

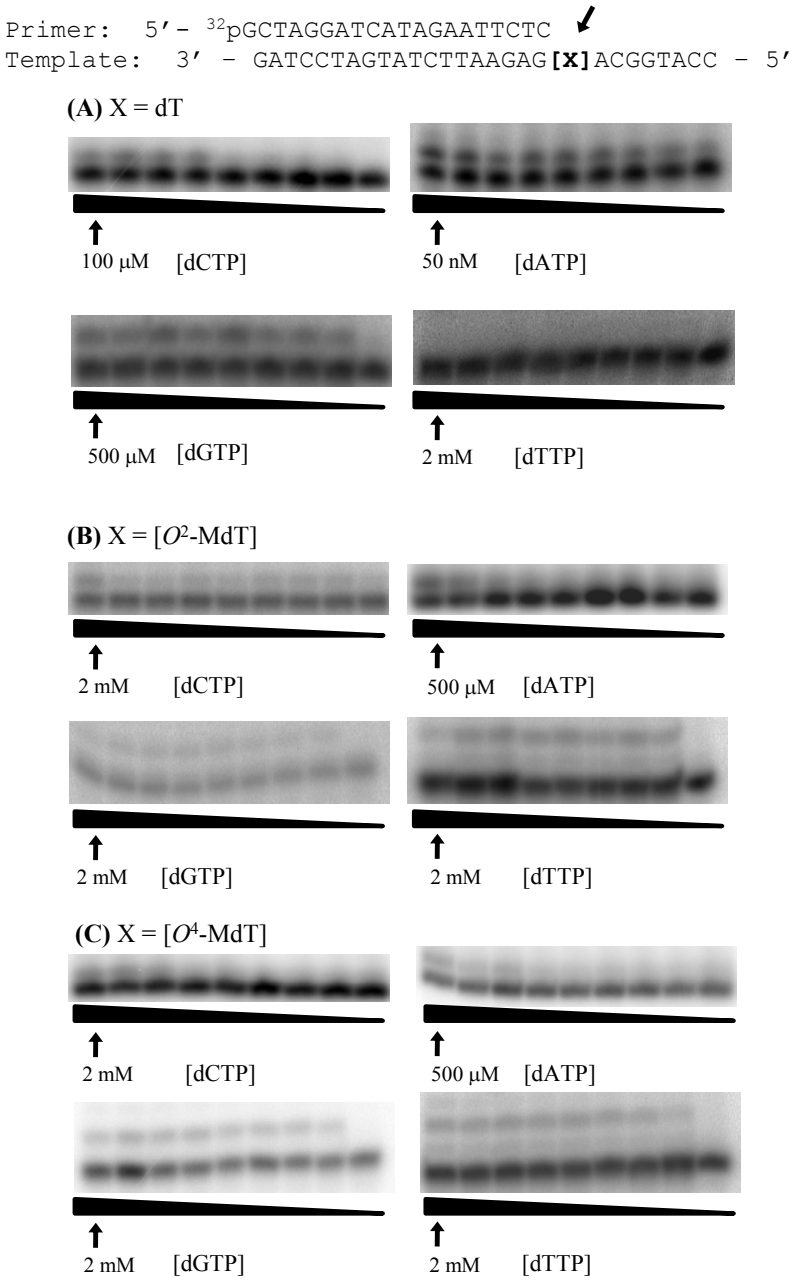


Figure A.5. Representative gel images for steady-state kinetic assays measuring nucleotide incorporation opposite undamaged dT (A), O^2-MdT (B), and O^4-MdT (C) with yeast polymerase η (7.5 nM). Reactions were conducted in the presence of individual dNTPs with the highest concentrations indicated in the figures. The concentration ratios of dNTP between neighboring lanes were 0.5-0.6.

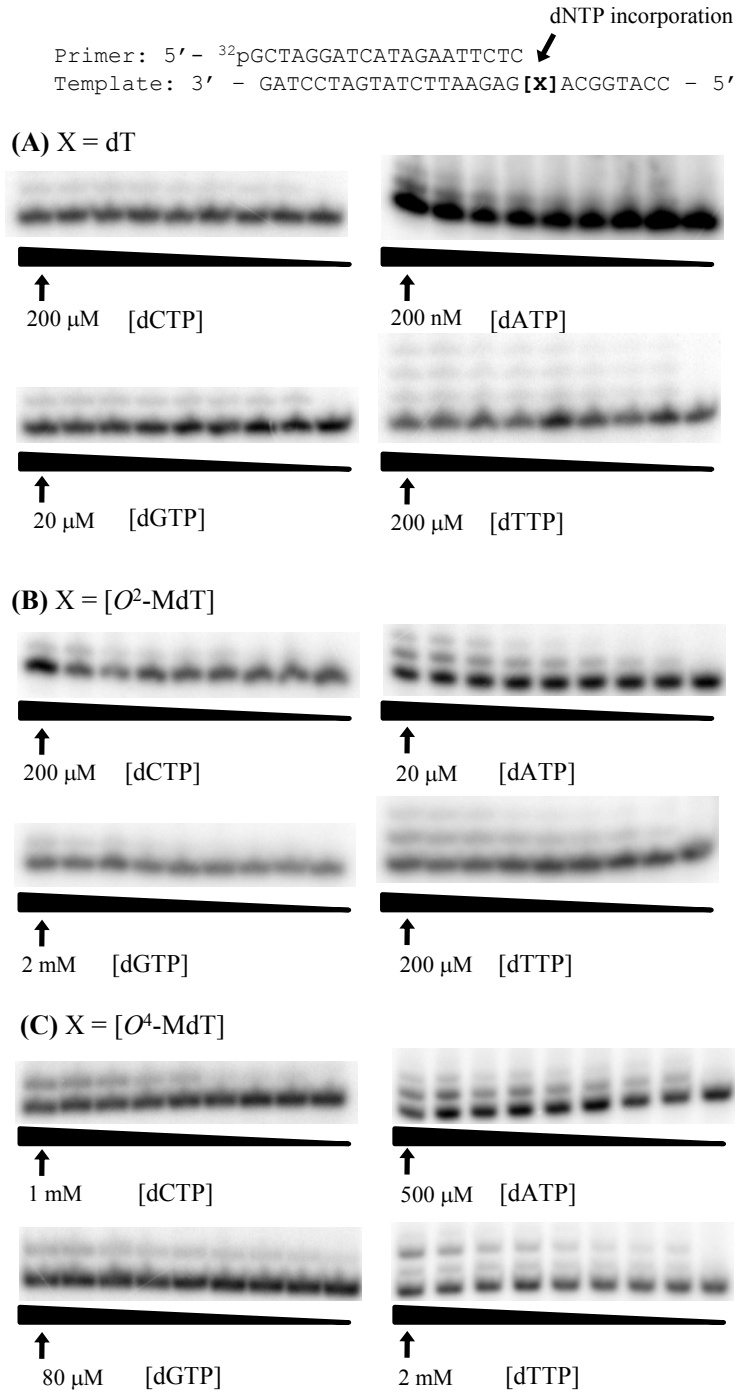
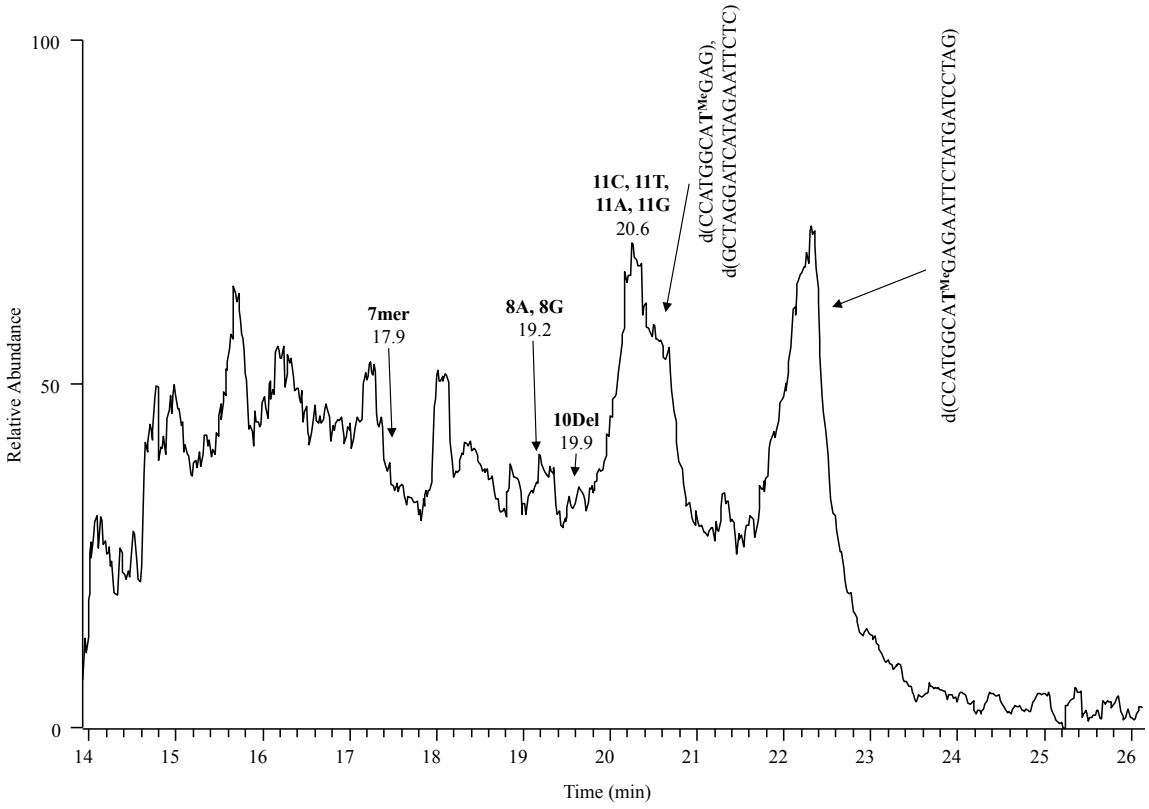


Figure A.6. (A) The total-ion chromatogram derived from the LC-MS/MS analysis of the human pol κ-induced replication products of O^2-MdT-containing substrates that have been treated with two restriction enzymes, NcoI and EcoRI plus shrimp alkaline phosphatase. (B) ESI-MS averaged from the peaks eluting in 18.5-20.5 min in part (A), where 'T*' designates the remnant of the digested template containing the damaged site, i.e., d(CATGGCA*X*GAG), where '*X*' is O^2-MdT, and 'P*' designates the 5' portion of the digested primer, i.e., d(GCTAGGATCATAG).

(A)



(B)

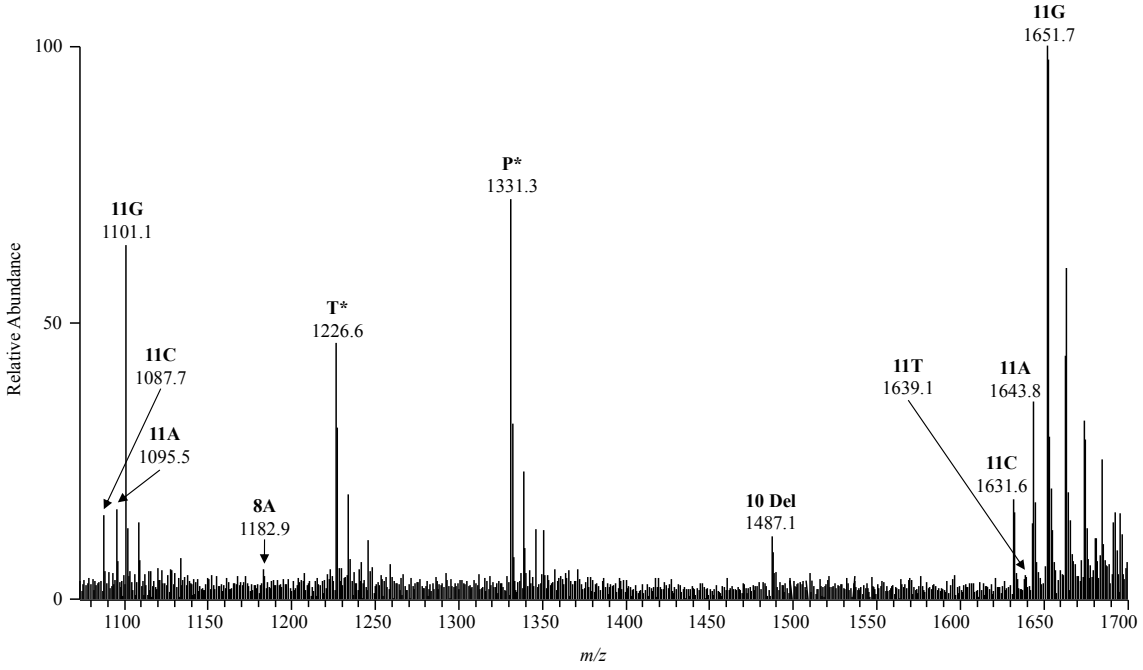
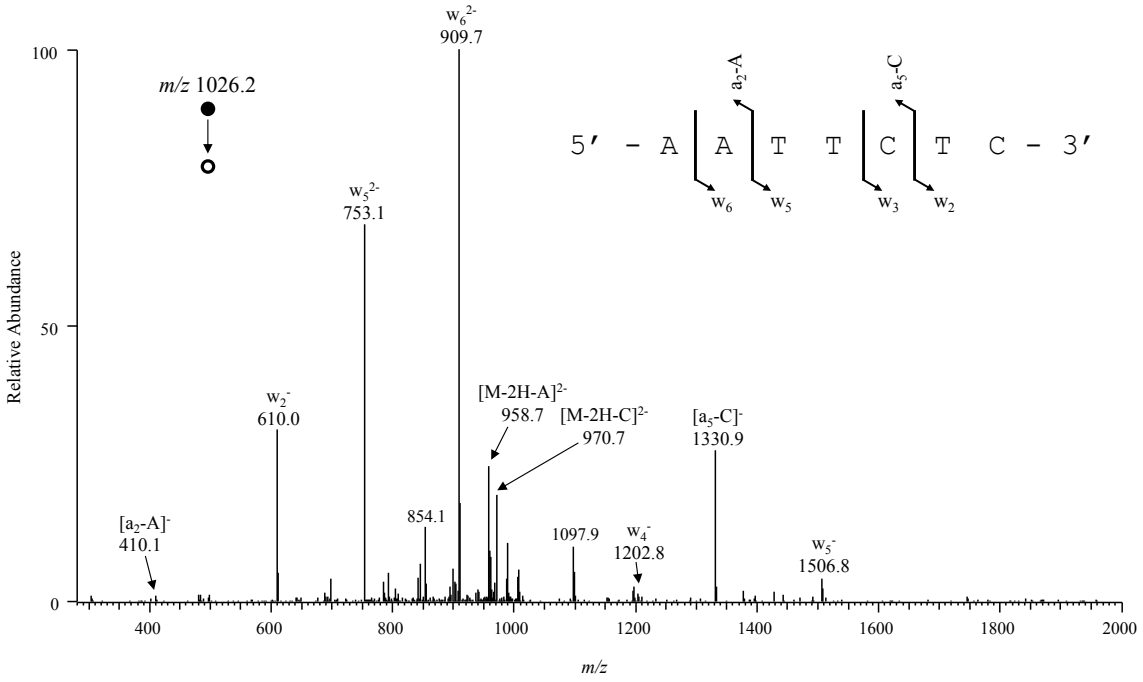
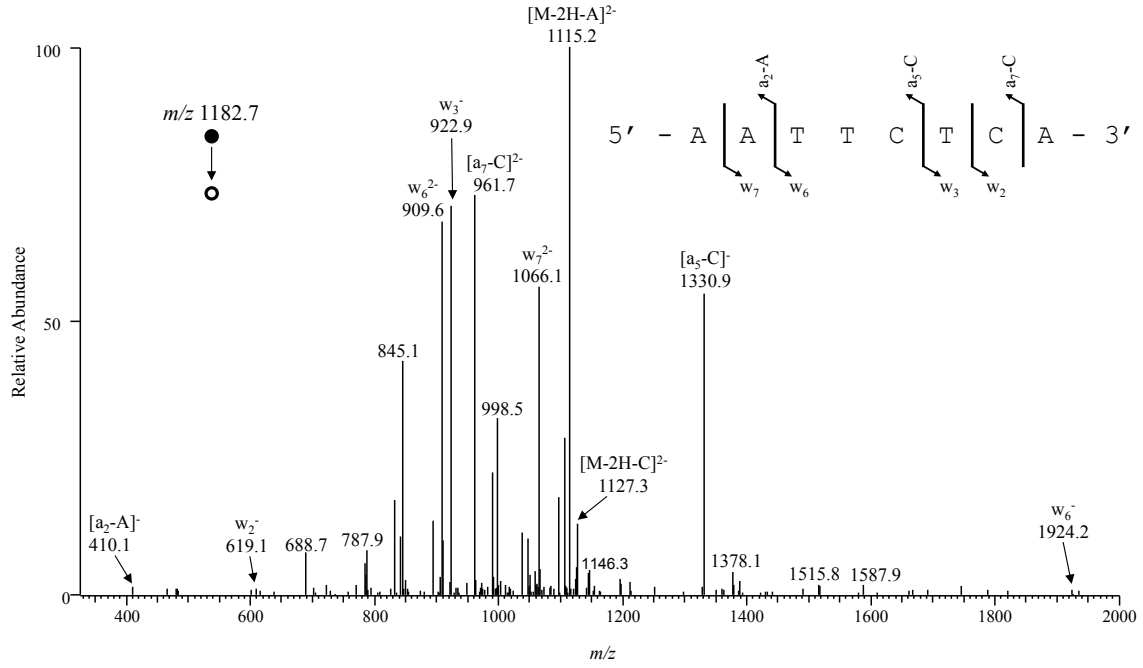

11G
1651.7
100
P*
1331.3
11G
1101.1
T*
1226.6
50
11C
1087.7
11T
1639.1
11A
1643.8
11A
1095.5
11C
1631.6
8A
1182.9
10 Del
1487.1
0
Relative Abundance
1100
1150
1200
1250
1300
1350
1400
1450
1500
1550
1600
1650
1700
m/z

Figure A.7. Analysis of the unextended primer and +1 extension products from the human pol κ-induced replication reaction arising from replication of O^2-MdT-containing substrates that have been treated with two restriction enzymes, NcoI and EcoRI, and shrimp alkaline phosphatase. Product-ion spectra of the ESI-produced $[M\text{-}2H]^{2-}$ ion of: the 7mer, d(AATTCTC) (precursor ion *m/z* 1026.2) (A); 8A, d(AATTCTCA) (precursor ion *m/z* 1182.7) (B); 8G, d(AATTCTCG) (precursor ion *m/z* 1190.7) (C); 10Del, d(AATTCTCTGC) (precursor ion *m/z* 1487.3) (D). In this and the following MS/MS figures, the arrow indicates the fragmentation process, and the solid and open circles represent the precursor ion and fragment ions, respectively.

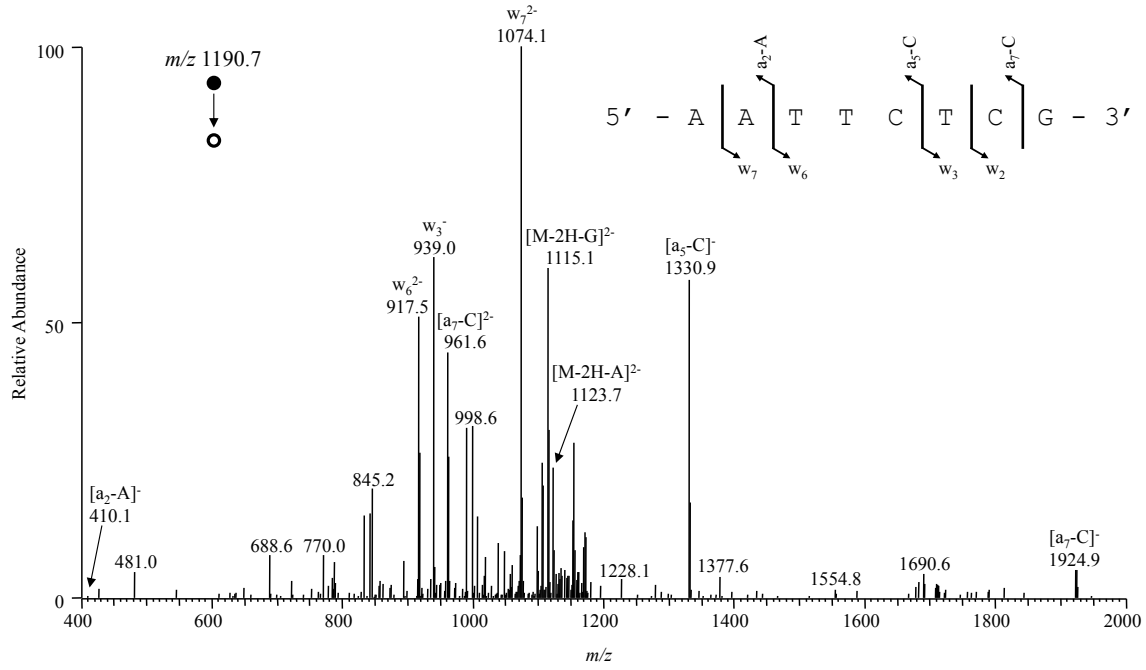
(A)



(B)


[M-2H-A]2-
1115.2
m/z 1182.7
w3-
922.9
[a7-C]2-
961.7
w62-
909.6
w72-
1066.1
[a5-C]-
1330.9
845.1
998.5
[M-2H-C]2-
1127.3
[a2-A]-
410.1
w2-
619.1
688.7
787.9
1146.3
1378.1
1515.8
1587.9
w6-
1924.2
5′ - A A T T C T C A - 3′
a2-A
a5-C
a7-C
w7
w6
w3
w2
Relative Abundance
m/z

(C)


w72-
1074.1
m/z 1190.7
w3-
939.0
[M-2H-G]2-
1115.1
[a5-C]-
1330.9
w62-
917.5
[a7-C]2-
961.6
[M-2H-A]2-
1123.7
998.6
845.2
[a2-A]-
410.1
481.0
688.6
770.0
1228.1
1377.6
1554.8
1690.6
[a7-C]-
1924.9
5′ - A A T T C T C G - 3′
a2-A
a5-C
a7-C
w7
w6
w3
w2
Relative Abundance
m/z

(D)

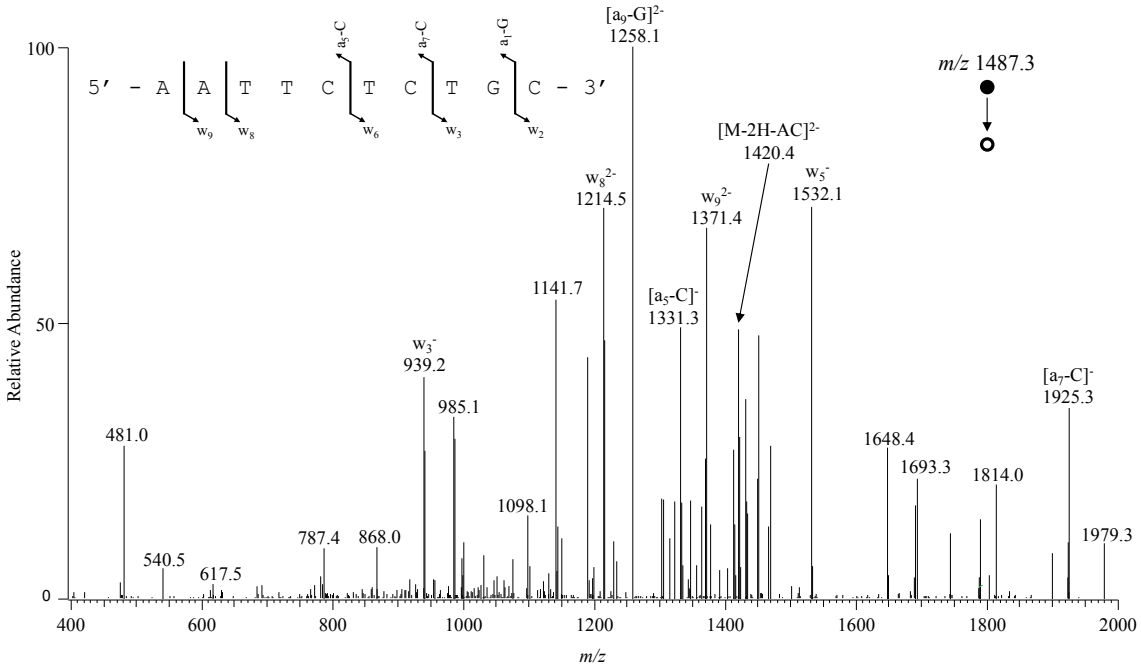
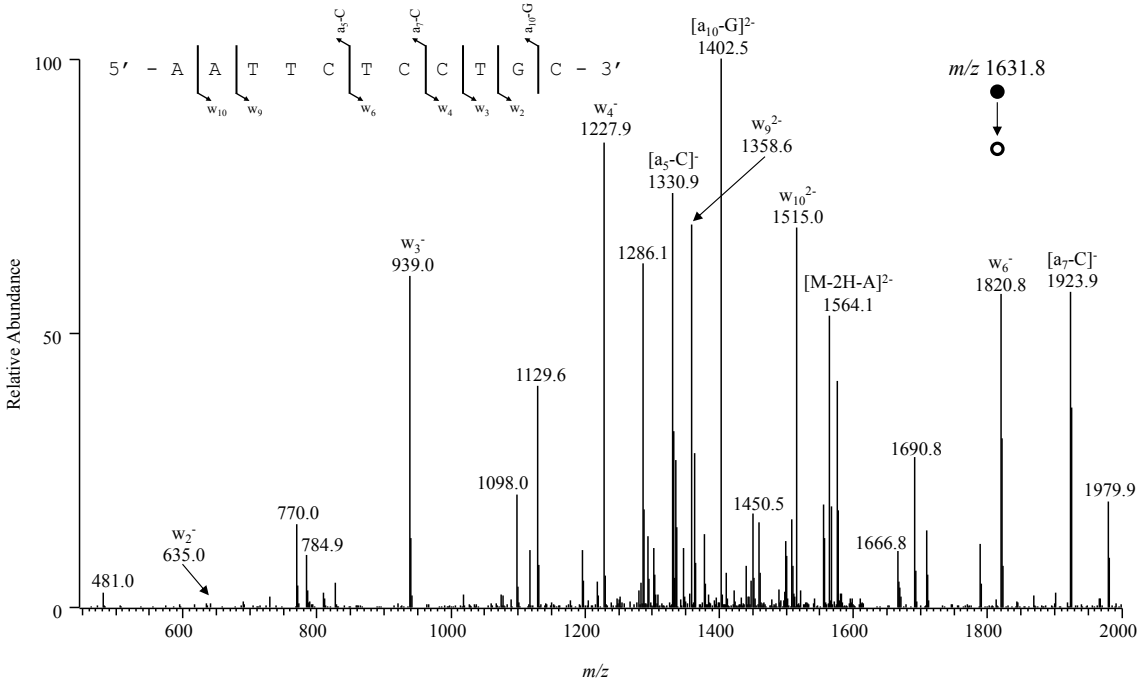
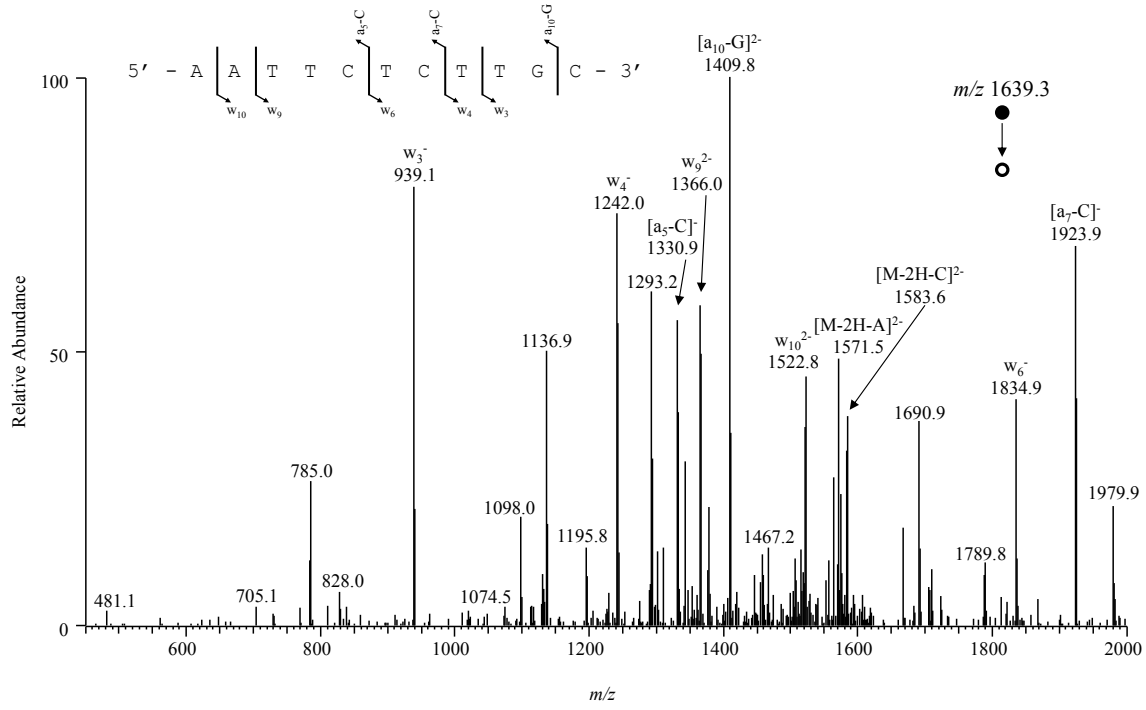

5′ - A A T T C T C T G C - 3′
w9
w8
w6
w3
w2
a5-C
a7-C
a1-G
[a0-G]2-
1258.1
m/z 1487.3
[M-2H-AC]2-
1420.4
w82-
1214.5
w92-
1371.4
w5-
1532.1
1141.7
[a5-C]-
1331.3
w3-
939.2
985.1
[a7-C]-
1925.3
481.0
1648.4
1693.3
1814.0
1098.1
787.4
868.0
1979.3
540.5
617.5
Relative Abundance
100
50
0
400
600
800
1000
1200
1400
1600
1800
2000
m/z

Figure A.8. Analysis of full-length extension products from the human pol κ-induced replication reaction arising from replication of O^2-MdT-containing substrates that have been treated with two restriction enzymes, NcoI and EcoRI, and shrimp alkaline phosphatase. Product-ion spectra of the ESI-produced $[M-2H]^{2-}$ ions of: 11C, d(AATTCTCCTGC) (precursor ion *m/z* 1631.8) (A); 11T, d(AATTCTCTTGC) (precursor ion *m/z* 1639.3) (B); 11A, d(AATTCTCATGC) (precursor ion *m/z* 1643.8) (C); 11G, d(AATTCTCGTGC) (precursor ion *m/z* 1651.7) (D).

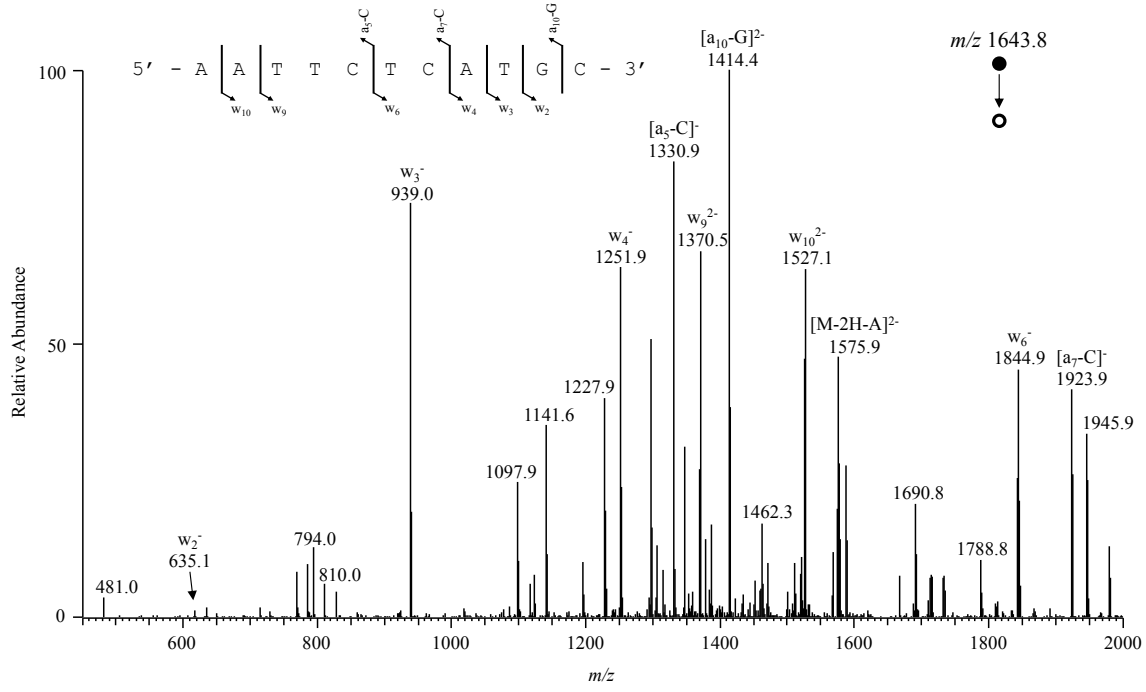
(A)



(B)


5′ - A A T T C T C T T G C - 3′
m/z 1639.3
[a10-G]2- 1409.8
w3- 939.1
w4- 1242.0
w92- 1366.0
[a5-C]- 1330.9
[a7-C]- 1923.9
1293.2
[M-2H-C]2- 1583.6
[M-2H-A]2- 1571.5
w102- 1522.8
1136.9
w6- 1834.9
1690.9
785.0
1979.9
1098.0
1195.8
1467.2
1789.8
828.0
481.1
705.1
1074.5
Relative Abundance
m/z

(C)


5′ - A A T T C T C A T G C - 3′
m/z 1643.8
[a10-G]2- 1414.4
[a5-C]- 1330.9
w3- 939.0
w92- 1370.5
w4- 1251.9
w102- 1527.1
[M-2H-A]2- 1575.9
w6- 1844.9
[a7-C]- 1923.9
1227.9
1141.6
1945.9
1097.9
1690.8
1462.3
w2- 635.1
794.0
1788.8
810.0
481.0
Relative Abundance
m/z

(D)

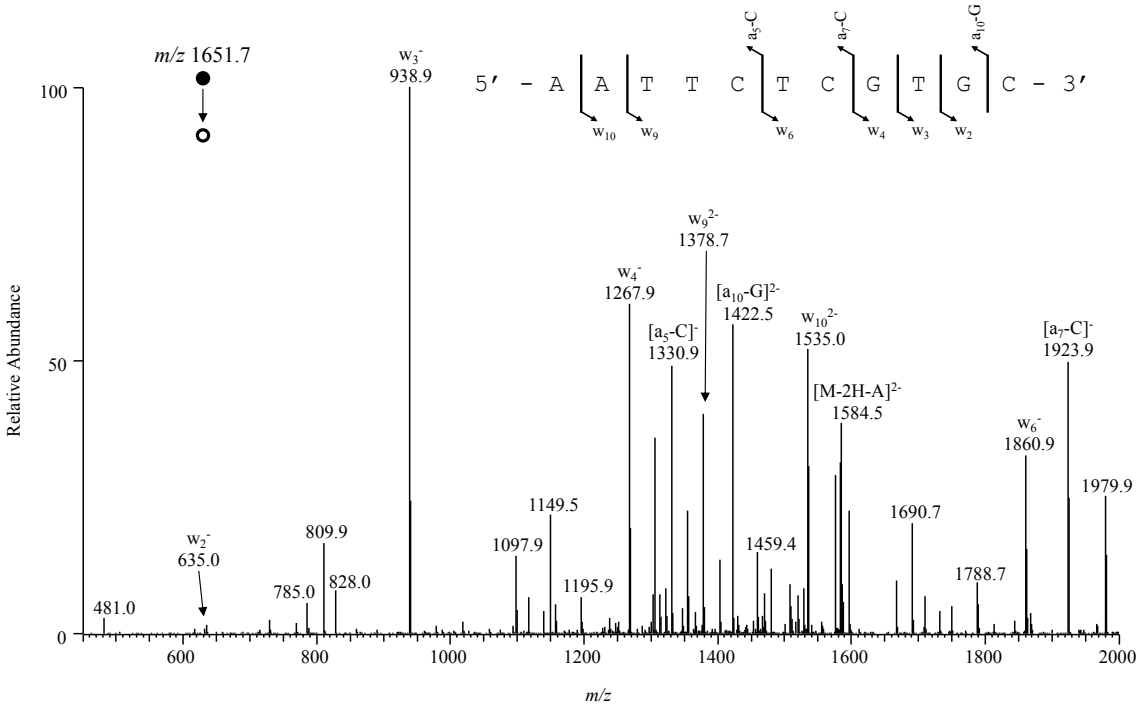

m/z 1651.7
w3- 938.9
5′ - A A T T C T C G T G C - 3′
a5-C
a7-C
a10-G
w10
w9
w6
w4
w3
w2
w9 2- 1378.7
w4- 1267.9
[a10-G]2- 1422.5
[a5-C]- 1330.9
w10 2- 1535.0
[a7-C]- 1923.9
[M-2H-A]2- 1584.5
w6- 1860.9
1979.9
1149.5
1690.7
809.9
w2- 635.0
1097.9
1459.4
1788.7
828.0
785.0
1195.9
481.0
Relative Abundance
100
50
0
600
800
1000
1200
1400
1600
1800
2000
m/z

Figure A.9. LC-MS/MS for monitoring possible demethylation of O^2-MdT-containing template following primer extension with human pol κ and restriction digestion. Shown are the SICs derived from ultra-zoom scans monitoring the $[M-3H]^{3-}$ ions from the unmethylated and O^2-MdT 12mer arising from the digestion of the original O^2-MdT-containing template.

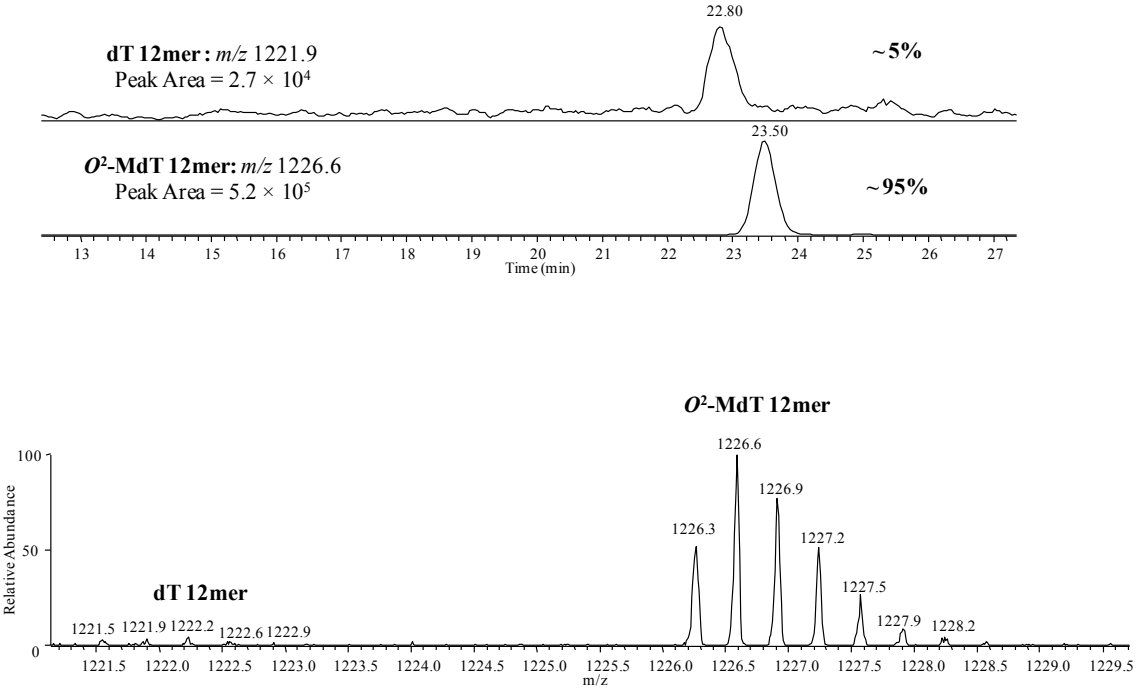
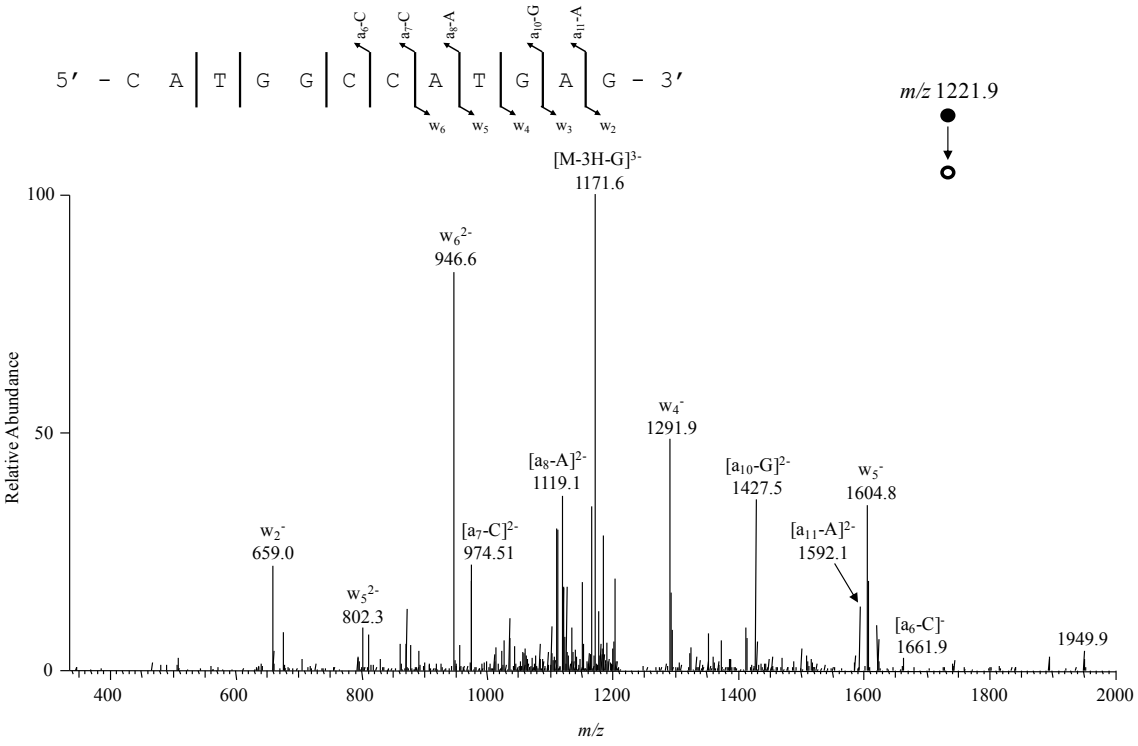


Figure A.10. LC-MS/MS for monitoring possible demethylation of O^2-MdT-containing templates following primer extension with human pol κ. Shown are the product-ion spectra of the ESI-produced $[M-3H]^{3-}$ ions of: unmethylated 12mer, d(CATGGCCA**T**GAG) (precursor ion *m/z* 1221.9) (A); original 12mer, d(CATGGCCA[***O*^{2}-MdT**]GAG) (precursor ion *m/z* 1226.5) (B).

(A)



(B)

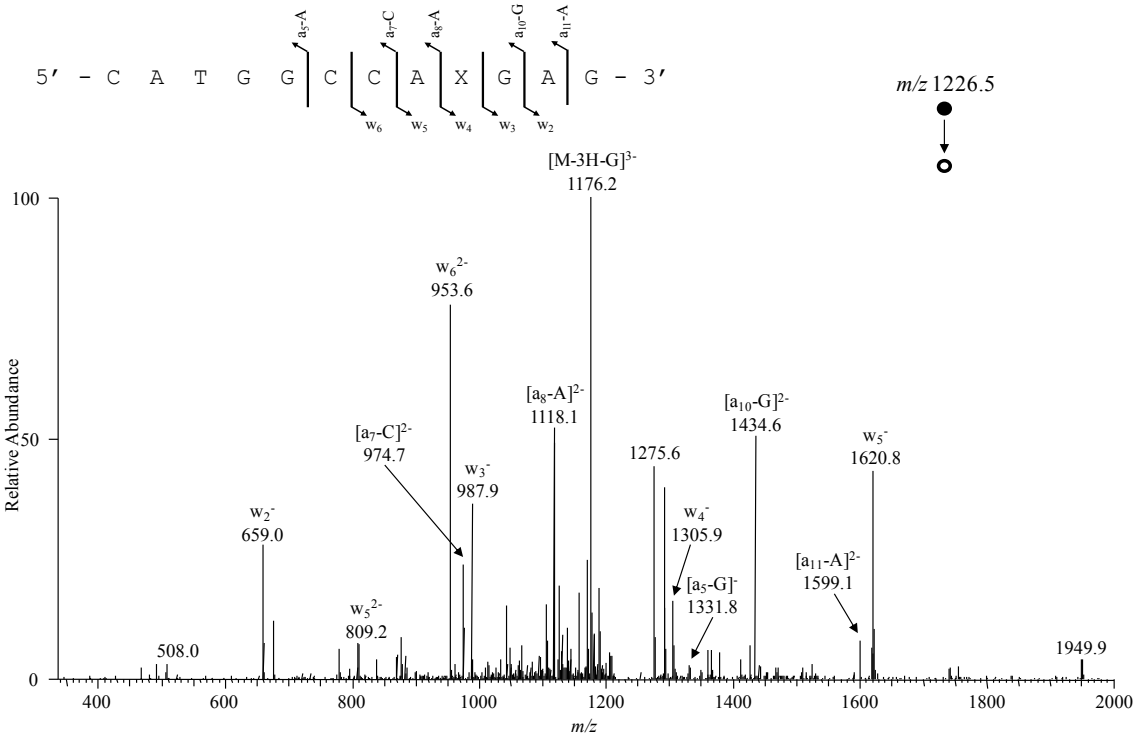

5′ - C A T G G C C A X G A G - 3′
a5-A
a7-C
a8-A
a10-G
a11-A
w6
w5
w4
w3
w2
m/z 1226.5
[M-3H-G]3-
1176.2
w6 2-
953.6
[a8-A]2-
1118.1
[a10-G]2-
1434.6
[a7-C]2-
974.7
w3-
987.9
1275.6
w5-
1620.8
w2-
659.0
w4-
1305.9
[a11-A]2-
1599.1
[a5-G]-
1331.8
w5 2-
809.2
508.0
1949.9
Relative Abundance
100
50
0
400
600
800
1000
1200
1400
1600
1800
2000
m/z

Figure A.11. Representative gel images for steady-state kinetic assays measuring extension past undamaged dT (A), O^2-MdT (B), and O^4-MdT (C) with the base opposite the lesion indicated in the figure. Reactions were conducted with human polymerase κ (1-5 nM) in the presence of individual dNTPs with the highest concentrations indicated in the figures. The concentration ratios of dNTP between neighboring lanes were 0.5-0.6.

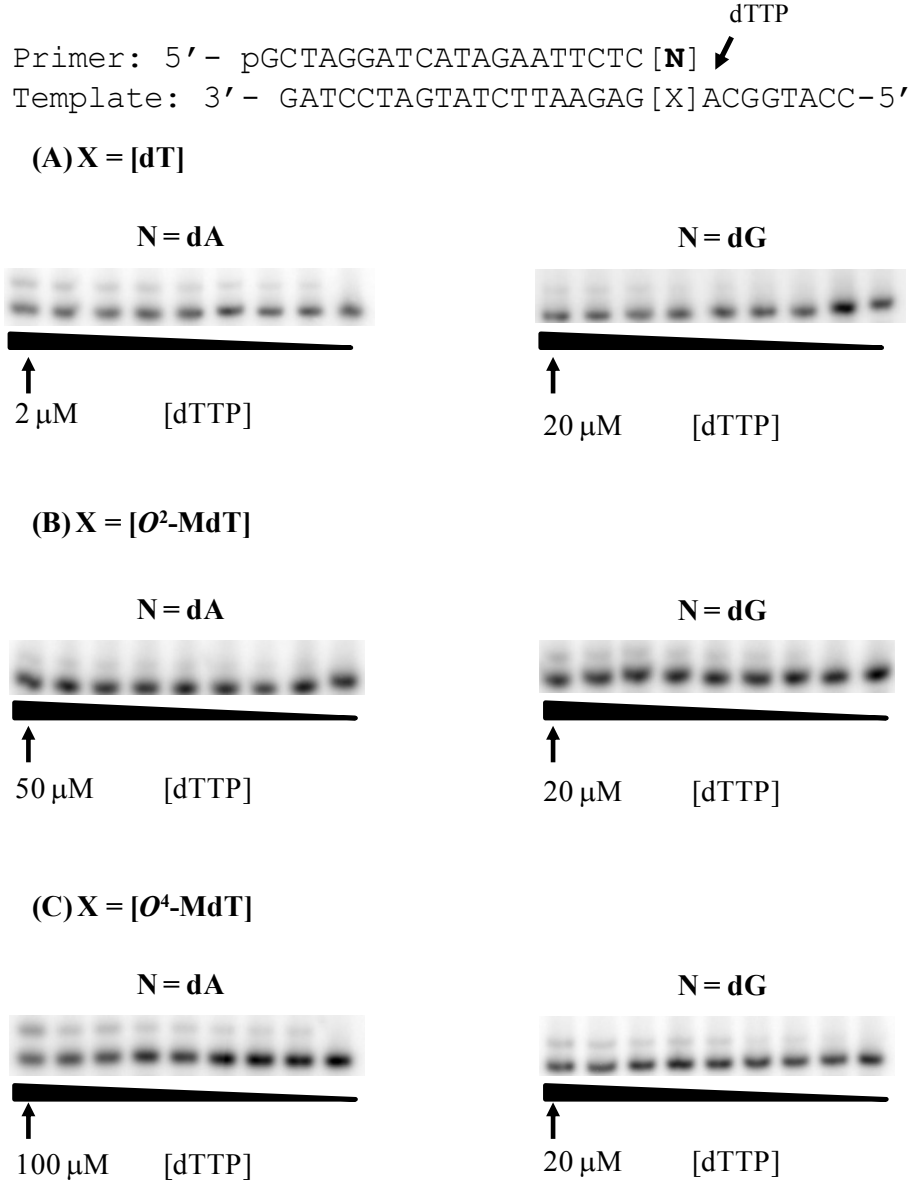


Table A.1. Summary of the percentages of replication products produced for O^4-MdT-containing substrates by yeast pol η as determined by LC-ESI-MS/MS experiments. Reactions were conducted under two separate conditions: (1) low dNTPs concentrations (50 μM each) at 37°C overnight; (2) 1 mM dNTPs at 37°C for 6 hr. The template was d(CCATGGCA***X***GAGAATTCTATGATCCTAG), where **'*X*'** represents O^4-MdT.

Name	Sequence	O^4-MdT (low dNTPs)	O^4-MdT (6 hr)
	Yeast Polymerase η		
7mer	d(AATTCTC)	6	26
8mer	d(AATTCTCA)	3	3
8mer	d(AATTCTCG)	8	7
11A	d(AATTCTCATGC)	10	3
11G	d(AATTCTCGTGC)	73	62

Table A.2. Efficiency and fidelity of human polymerase κ-mediated nucleotide extension with dATP and dGTP opposite undamaged dT, O^2-MdT and O^4-MdT as determined by steady-state kinetic measurements.

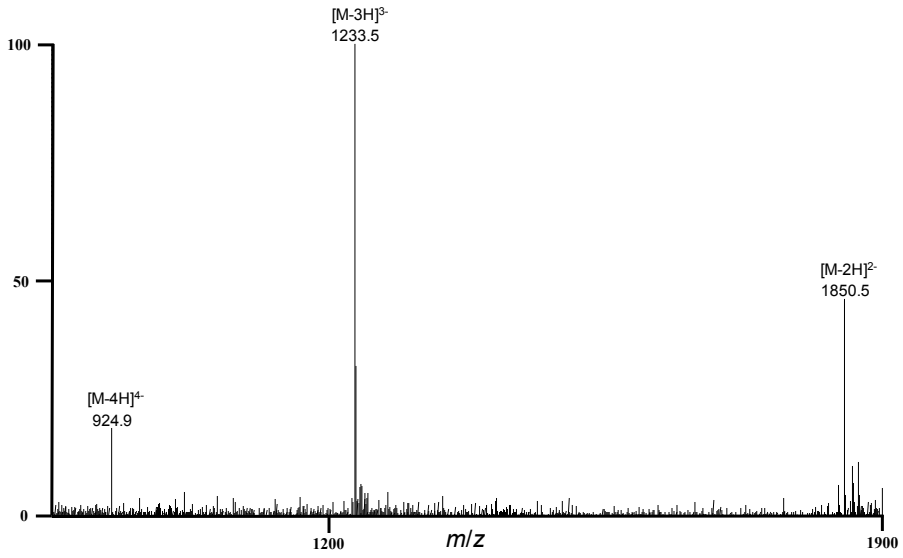
dNTP	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (mM^{-1} min^{-1})	f_{ext}
		Extension with dTTP		
dT:dATP	9.5 ± 0.4	0.56 ± 0.02	17	1
dT:dGTP	5.3 ± 0.1	2.2 ± 0.3	2.5	0.15
O^2-MdT:dATP	9.5 ± 0.4	7.8 ± 0.8	1.2	1
O^2-MdT:dGTP	13 ± 1	7.5 ± 0.5	1.7	1.4
O^4-MdT:dATP	10 ± 1	26 ± 2	0.38	1
O^4-MdT:dGTP	27 ± 2	11 ± 1	2	5.2

*The K_m and k_{cat} were average values based on three independent measurements

Appendix B: Supporting Information for Chapter 3

Figure B.1. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^2-EtdT Negative-ion ESI-MS (other peaks are HFIP adducts) (A); the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1233.5) (B).

(A)



(B)

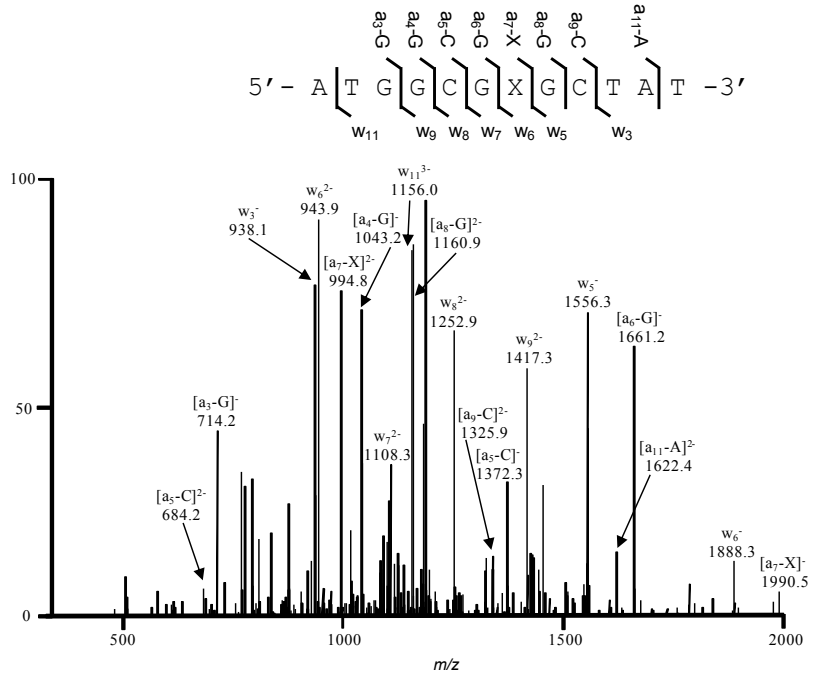
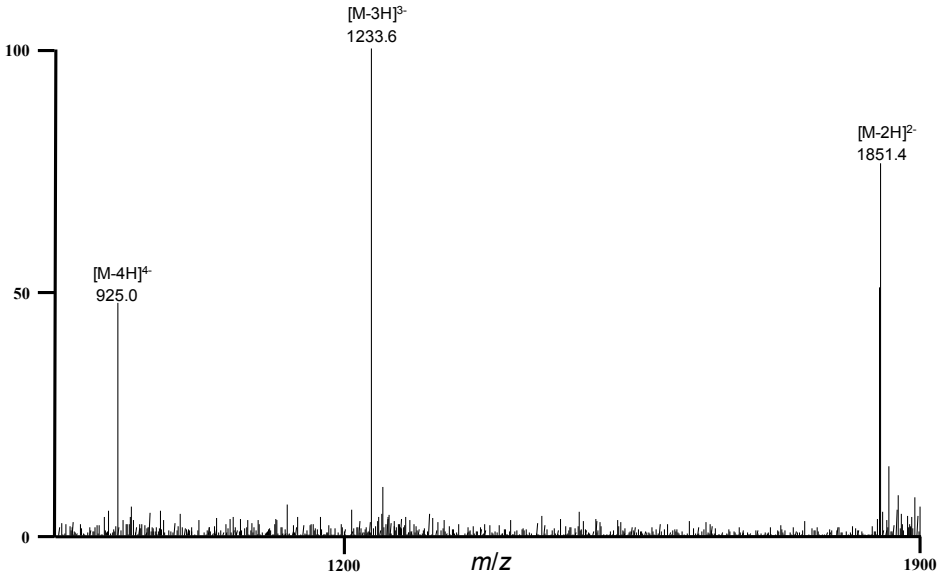


Figure B.2. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=*N*3-EtdT Negative-ion ESI-MS (other peaks are HFIP adducts) (A); the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1233.6) (B).

(A)



(B)

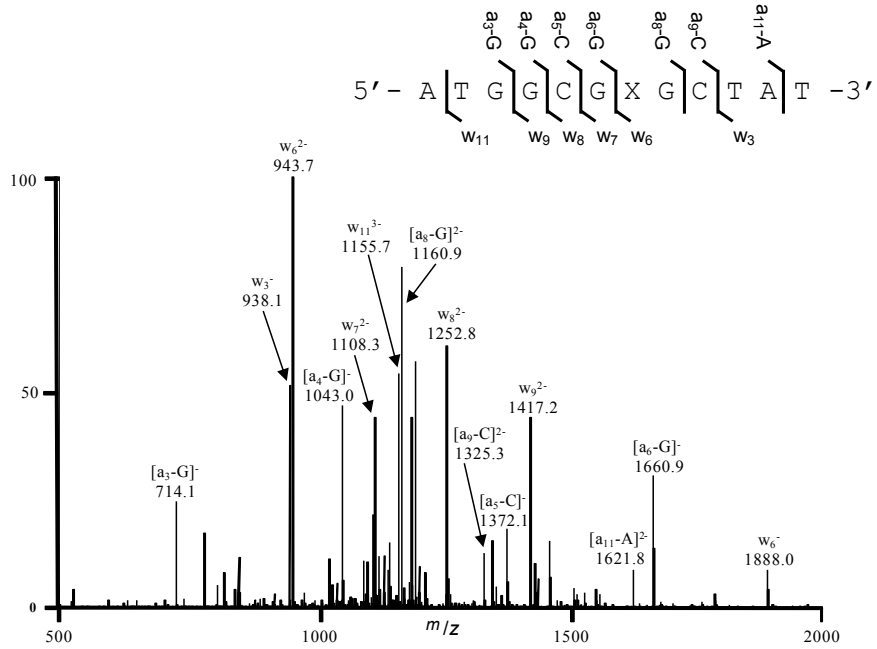
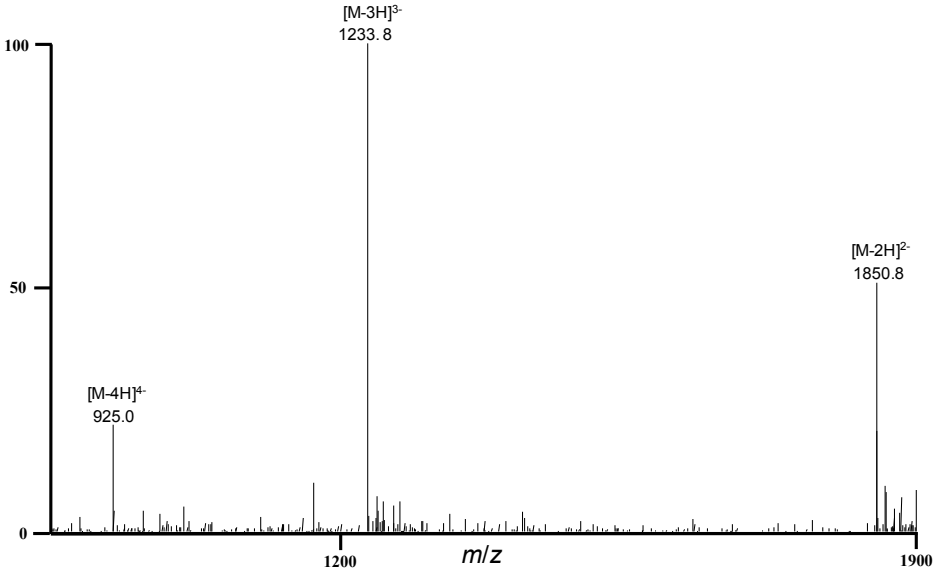


Figure B.3. ESI-MS & MS/MS characterizations of d(ATGGCGXGCTAT), X=O^4-EtdT Negative-ion ESI-MS (other peaks are HFIP adducts) (A); the product-ion spectrum of the $[M-3H]^{3-}$ ion (*m*/*z* 1233.7) (B).

(A)



(B)

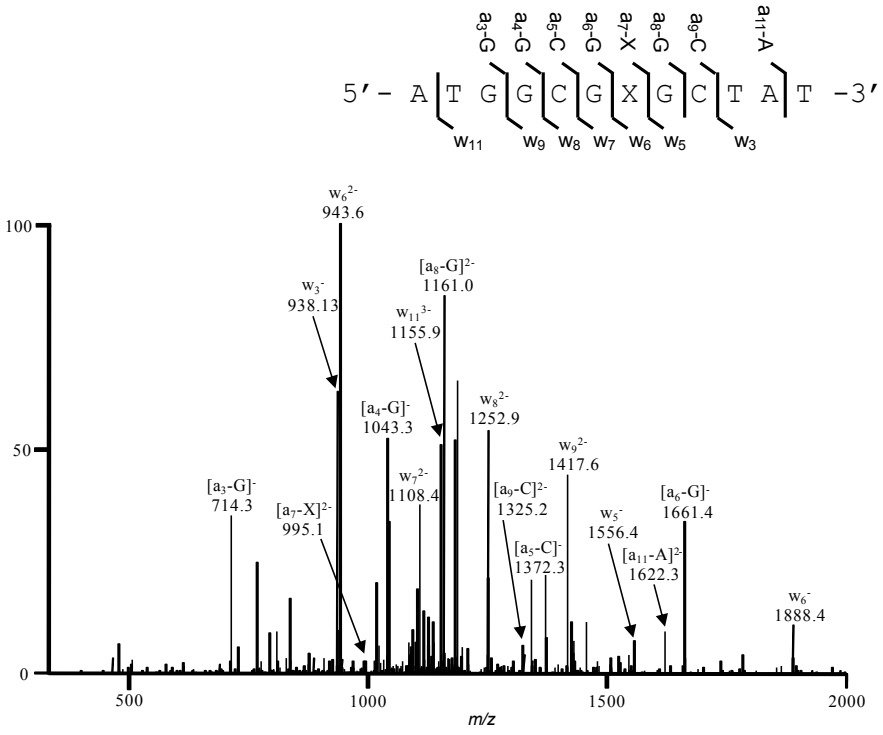
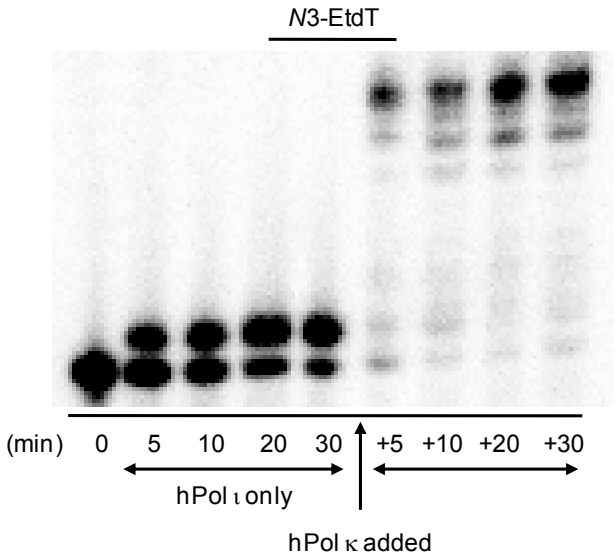


Figure B.4. Cooperativity of human DNA polymerase ι and κ during the bypass of *N*3-EtdT (A) and O^4-EtdT lesion (B). Primer-template complex containing *N*3-EtdT and O^4-EtdT (50 nm) was incubated with hPol ι (10 nM) for 0–30 min. After 30 min, 5 nM of hPol κ was added to the reaction mixture to determine whether the primer extension started by hPol ι and stalled by the lesion can be completed by hPol κ. The extension products were resolved by PAGE.

(A)



(B)

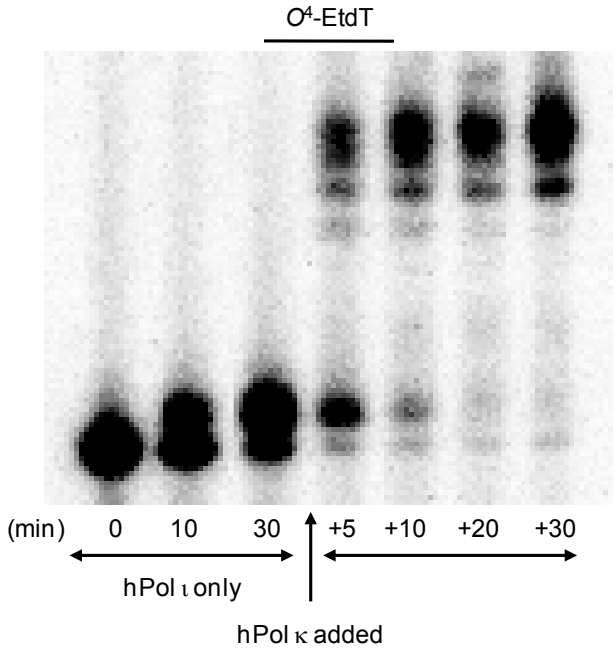
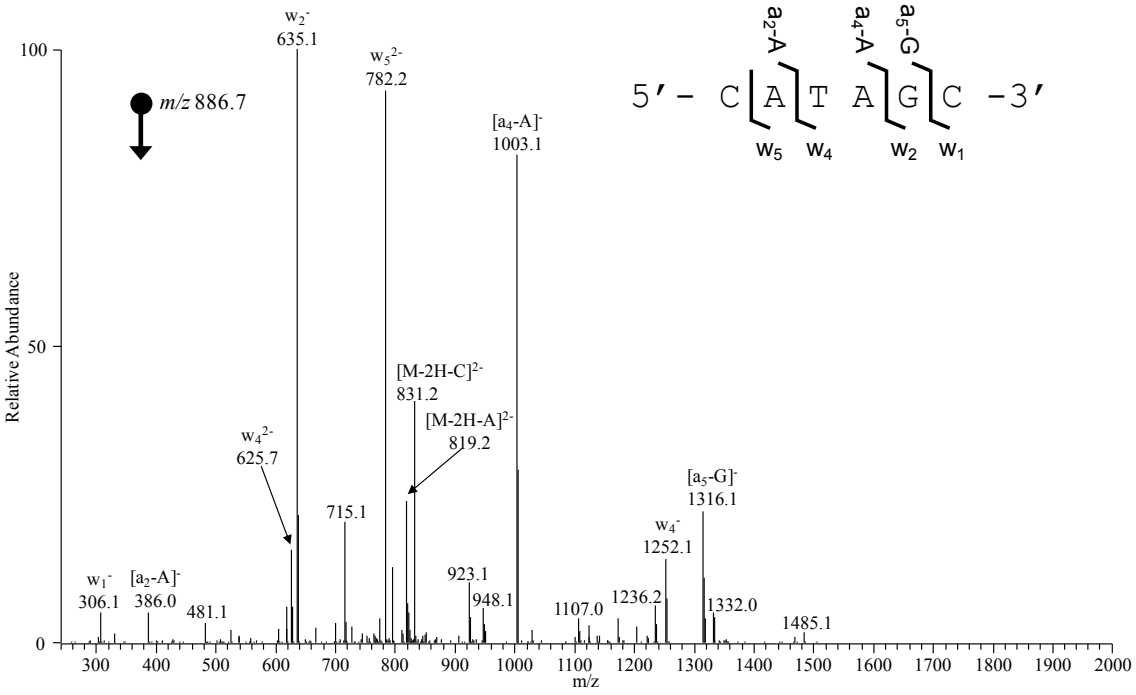
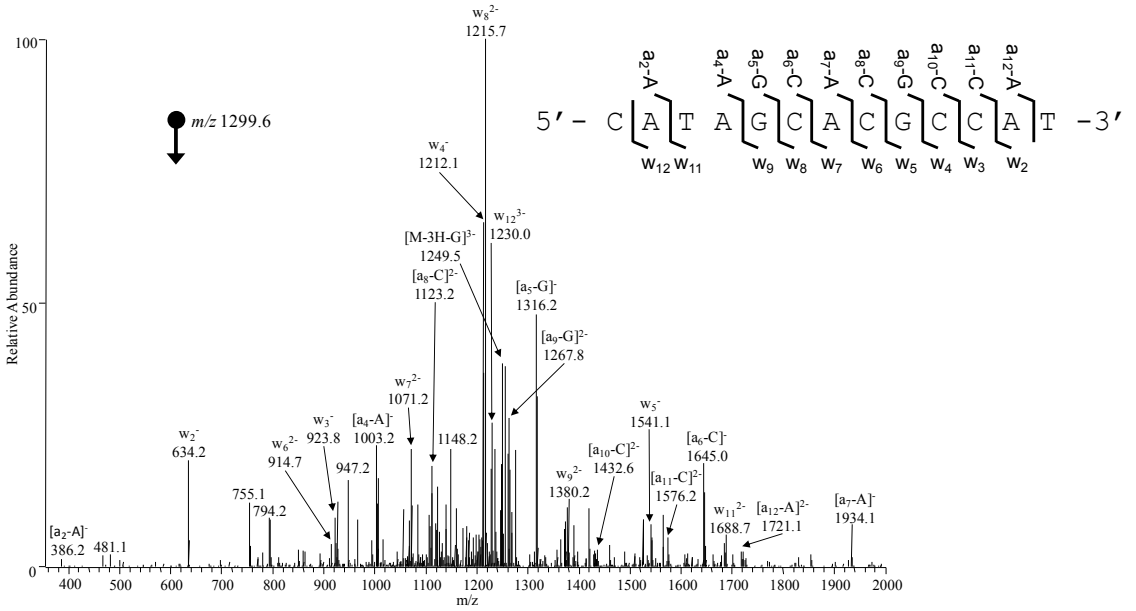


Figure B.5. Analysis of the unextended primer and extension products from hPol η-induced replication reaction arising of O^2-EtdT-containing substrate that had been treated with the restriction enzyme SacI and shrimp alkaline phosphatase. The product-ion spectra of the ESI-produced $[M-2H]^{2-}$ ion of: the 6mer, d(CATAGC) (precursor ion *m/z* 886.7) (A). The product-ion spectra of the ESI-produced $[M-3H]^{3-}$ ion of: 13A, d(CATAGCACGCCAT) (precursor ion *m/z* 1299.6) (B); 13G, d(CATAGCGCGCCAT) (precursor ion *m/z* 1304.9) (C); 14AA, d(CATAGCACGCCATA) (precursor ion *m/z* 1403.9) (D); 14GA, d(CATAGCGCGCCATA) (precursor ion *m/z* 1409.2) (E). In this figure the arrow indicates the fragmentation process, and the solid and open circles represent the precursor ion and fragment ions, respectively.

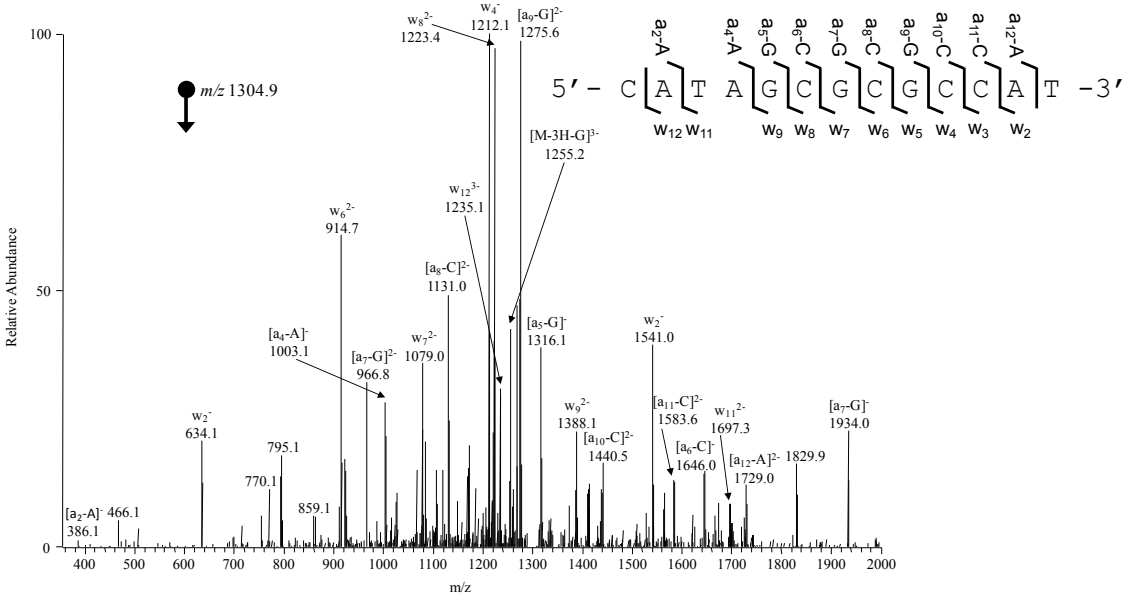
(A)



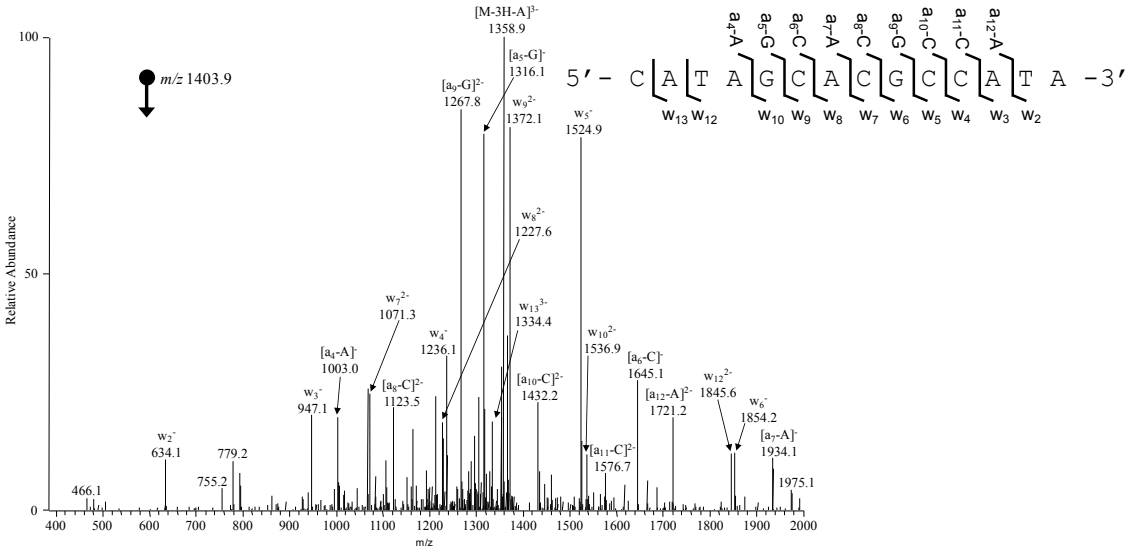
(B)


m/z 1299.6
5′- C A T A G C A C G C C A T -3′
Relative Abundance
m/z
w8 2- 1215.7
w4- 1212.1
w12 3- 1230.0
[M-3H-G]3- 1249.5
[a8-C]2- 1123.2
[a5-G]- 1316.2
[a9-G]2- 1267.8
w7 2- 1071.2
w5- 1541.1
w3- 923.8
[a4-A]- 1003.2
w2- 634.2
w6 2- 914.7
1148.2
[a6-C]- 1645.0
[a10-C]2- 1432.6
947.2
w9 2- 1380.2
[a11-C]2- 1576.2
755.1
794.2
w11 2- 1688.7
[a12-A]2- 1721.1
[a7-A]- 1934.1
[a2-A]- 386.2
481.1

(C)


m/z 1304.9
5′- C A T A G C G C G C C A T -3′
Relative Abundance
m/z
w8 2- 1223.4
w4- 1212.1
[a9-G]2- 1275.6
[M-3H-G]3- 1255.2
w12 3- 1235.1
w6 2- 914.7
[a8-C]2- 1131.0
[a5-G]- 1316.1
w2- 1541.0
[a4-A]- 1003.1
[a7-G]2- 966.8
w7 2- 1079.0
w9 2- 1388.1
[a11-C]2- 1583.6
w11 2- 1697.3
[a7-G]- 1934.0
w2- 634.1
[a10-C]2- 1440.5
795.1
[a6-C]- 1646.0
[a12-A]2- 1729.0
1829.9
770.1
859.1
[a2-A]- 386.1
466.1

(D)


m/z 1403.9
5′- C A T A G C A C G C C A T A -3′
a4-A a5-G a6-C a7-A a8-C a9-G a10-C a11-C a12-A
w13 w12 w10 w9 w8 w7 w6 w5 w4 w3 w2
[M-3H-A]3- 1358.9
[a5-G]- 1316.1
[a9-G]2- 1267.8
w92- 1372.1
w5- 1524.9
w92- 1227.6
w72- 1071.3
w133- 1334.4
w4- 1236.1
w102- 1536.9
[a6-C]- 1645.1
[a4-A]- 1003.0
[a8-C]2- 1123.5
[a10-C]2- 1432.2
w122- 1845.6
w3- 947.1
[a12-A]2- 1721.2
w6- 1854.2
w2- 634.1
[a11-C]2- 1576.7
[a7-A]- 1934.1
779.2
755.2
466.1
1975.1
Relative Abundance
m/z

(E)

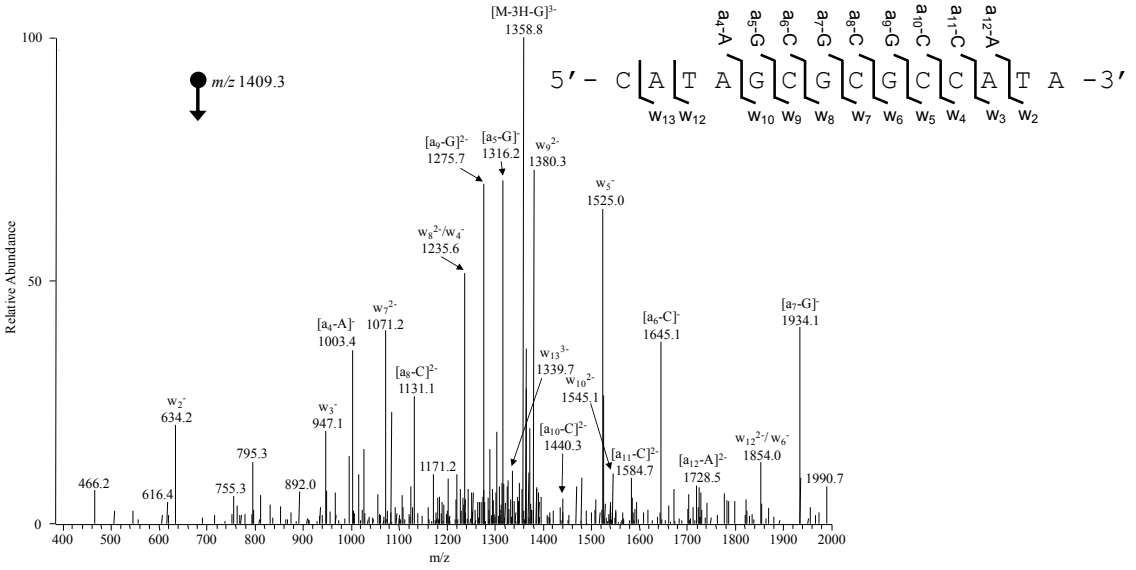
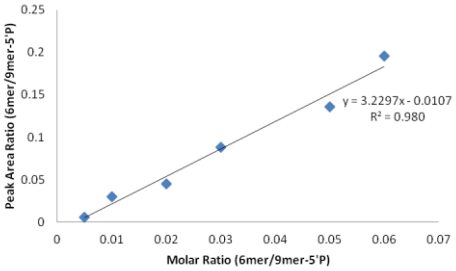
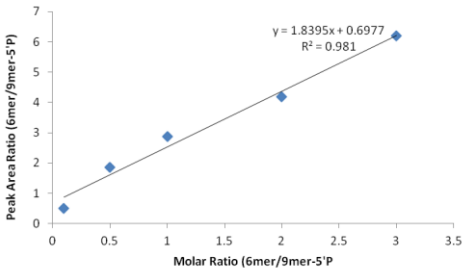

m/z 1409.3
5′- C A T A G C G C G C C A T A -3′
a4-A a5-G a6-C a7-G a8-C a9-G a10-C a11-C a12-A
w13 w12 w10 w9 w8 w7 w6 w5 w4 w3 w2
[M-3H-G]3- 1358.8
[a9-G]2- 1275.7
[a5-G]- 1316.2
w92- 1380.3
w5- 1525.0
w82-/w4- 1235.6
[a7-G]- 1934.1
[a6-C]- 1645.1
w72- 1071.2
[a4-A]- 1003.4
w133- 1339.7
w102- 1545.1
[a8-C]2- 1131.1
w2- 634.2
w3- 947.1
[a10-C]2- 1440.3
[a11-C]2- 1584.7
w122-/w6- 1854.0
[a12-A]2- 1728.5
795.3
1171.2
466.2
616.4
755.3
892.0
1990.7
Relative Abundance
m/z

Figure B.6. Calibration curves for the unextended primer d(CATAGC) (i.e., 6mer) to d(GATCGAGCT) (i.e., 9-mer 5'P) (A); d(CATAGC) (i.e., 6mer) to 9-mer 5'P (B); d(CATAGCC) (i.e., 7C) to 9-mer 5'P (C); d(CATAGCT) (i.e., 7T) to 9-mer 5'P (D); d(CATAGCA) (i.e., 7A) to 9-mer 5'P (E); d(CATAGCG) (i.e, 7G) to 9-mer 5'P (F); d(CATAGCAC) (i.e., 8A) to 9-mer 5'P (G); d(CATAGCGC) (i.e., 8G) to 9-mer 5'P (H); d(CATAGCCCG) (i.e., 9C) to 9-mer 5'P (I); d(CATAGCACG) (i.e., 9A) to 9-mer 5'P (J); d(CATAGCGCG) (i.e., 9G) to 9-mer 5'P (K); d(CATAGCGCG) (i.e., 9G) to 9-mer 5'P (L); d(CATAGCCGC) (i.e., 10Del @ lesion site) to 9-mer 5'P (M); d(CATAGCAGCC) (i.e., 10Adel +1 from lesion site) to 9-mer 5'P (N); d(CATAGCCGCCA) (i.e., 11Del @ lesion site) to 9-mer 5'P (O); d(CATAGCACGCCA) (i.e., 12A) to 9-mer 5'P (P); d(CATAGCACGCCA) (i.e., 12A) to 9-mer 5'P (Q); d(CATAGCGCGCCA) (i.e., 12G) to 9-mer 5'P (R); d(CATAGCACGCCAT) (i.e., 13A) to 9-mer 5'P (S); d(CATAGCACGCCAT) (i.e., 13A) to 9-mer 5'P (T); d(CATAGCGCGCCAT) (i.e., 13G) to 9-mer 5'P (U); d(CATAGCGCGCCAT) (i.e., 13G) to 9-mer 5'P (V); d(CATAGCCGCCATA) (i.e., 14TA) to 9-mer 5'P (W); d(CATAGCACGCCATA) (i.e., 14AA) to 9-mer 5'P (X); d(CATAGCGCGCCATA) (i.e., 14GA) to 9-mer 5'P (Y); d(CATAGCGCGCCATA) (i.e., 14GA) to 9-mer 5'P (Z); d(CATAGCTCGCCATAT) (i.e., 15TAT) to 9-mer 5'P (AA); d(CATAGCACGCCATAT) (i.e., 15AAT) to 9-mer 5'P (BB); d(CATAGCACGCCATAT) (i.e., 15AAT) to 9-mer 5'P (CC); d(CATAGCGCGCCATAT) (i.e., 15GAT) to 9-mer 5'P (DD); d(CATAGCGCGCCATAT) (i.e., 15GAT) to 9-mer 5'P (EE). In some cases, two calibration curves were constructed for different concentration ranges so that the ODNs could be accurately quantified over a wide concentration range.

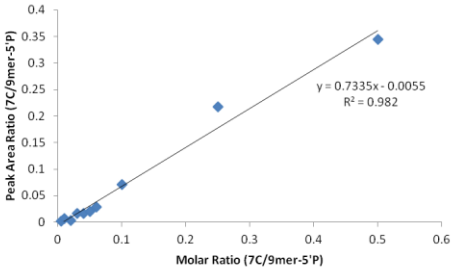
(A)


y = 3.2297x - 0.0107
R² = 0.980
Peak Area Ratio (6mer/9mer-5'P)
Molar Ratio (6mer/9mer-5'P)

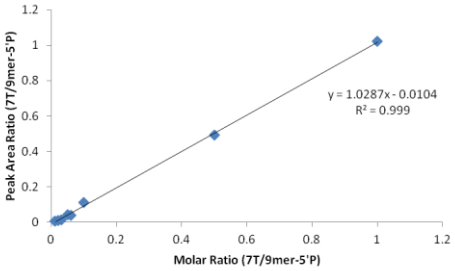
(B)


y = 1.8395x + 0.6977
R² = 0.981
Peak Area Ratio (6mer/9mer-5'P)
Molar Ratio (6mer/9mer-5'P

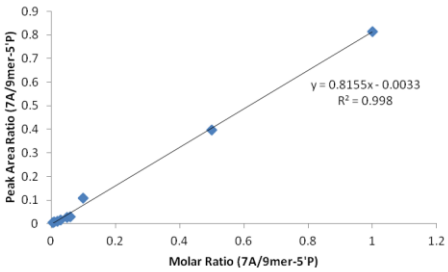
(C)


y = 0.7335x - 0.0055
R² = 0.982
Peak Area Ratio (7C/9mer-5'P)
Molar Ratio (7C/9mer-5'P)

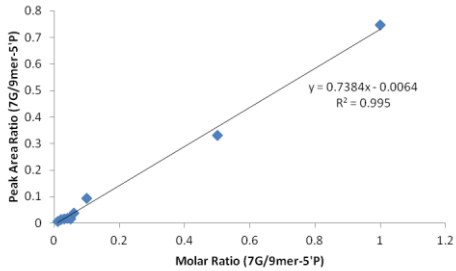
(D)


y = 1.0287x - 0.0104
R² = 0.999
Peak Area Ratio (7T/9mer-5'P)
Molar Ratio (7T/9mer-5'P)

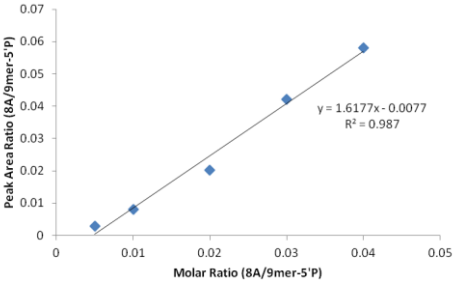
(E)


y = 0.8155x - 0.0033
R² = 0.998
Peak Area Ratio (7A/9mer-5'P)
Molar Ratio (7A/9mer-5'P)

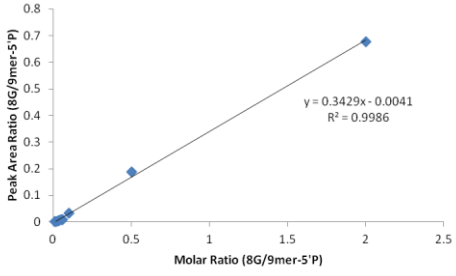
(F)


y = 0.7384x - 0.0064
R² = 0.995
Peak Area Ratio (7G/9mer-5'P)
Molar Ratio (7G/9mer-5'P)

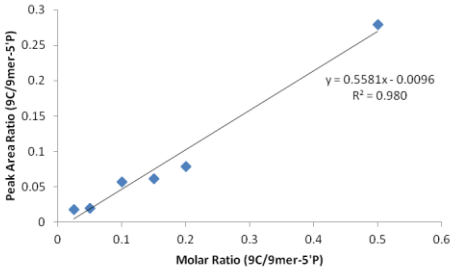
(G)


y = 1.6177x - 0.0077
R² = 0.987
Peak Area Ratio (8A/9mer-5'P)
Molar Ratio (8A/9mer-5'P)

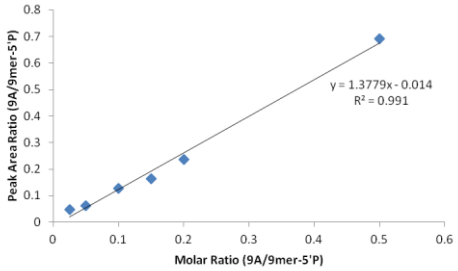
(H)


y = 0.3429x - 0.0041
R² = 0.9986
Peak Area Ratio (8G/9mer-5'P)
Molar Ratio (8G/9mer-5'P)

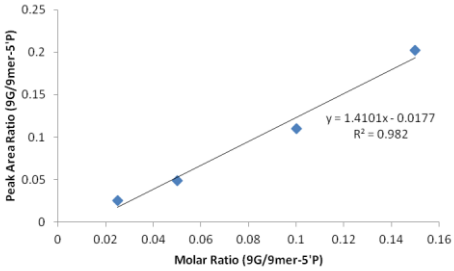
(I)


y = 0.5581x - 0.0096
R² = 0.980
Peak Area Ratio (9C/9mer-5'P)
Molar Ratio (9C/9mer-5'P)

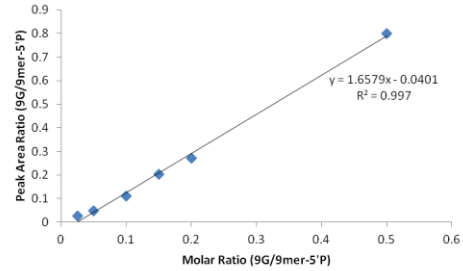
(J)


y = 1.3779x - 0.014
R² = 0.991
Peak Area Ratio (9A/9mer-5'P)
Molar Ratio (9A/9mer-5'P)

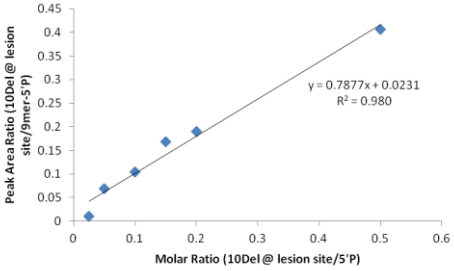
(K)


y = 1.4101x - 0.0177
R² = 0.982
Peak Area Ratio (9G/9mer-5'P)
Molar Ratio (9G/9mer-5'P)

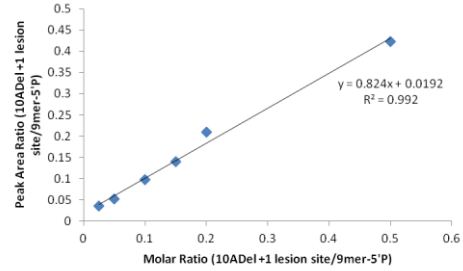
(L)


y = 1.6579x - 0.0401
R² = 0.997
Peak Area Ratio (9G/9mer-5'P)
Molar Ratio (9G/9mer-5'P)

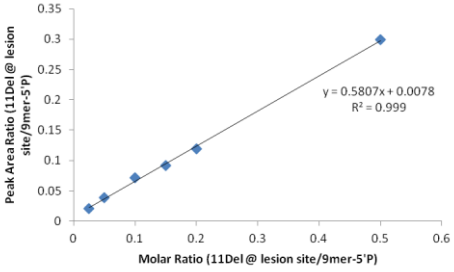
(M)


y = 0.7877x + 0.0231
R² = 0.980
Peak Area Ratio (10Del @ lesion site/9mer-5'P)
Molar Ratio (10Del @ lesion site/5'P)

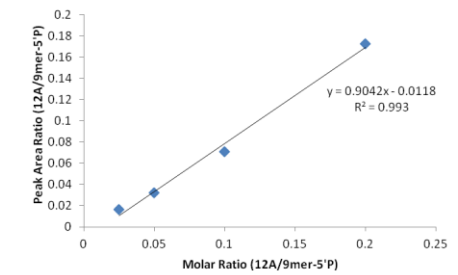
(N)


y = 0.824x + 0.0192
R² = 0.992
Peak Area Ratio (10ADel +1 lesion site/9mer-5'P)
Molar Ratio (10ADel +1 lesion site/9mer-5'P)

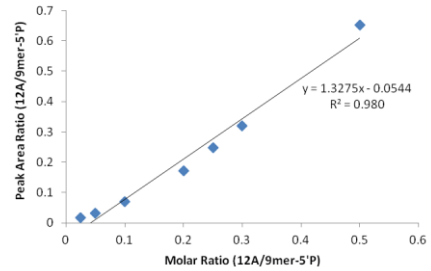
(O)


y = 0.5807x + 0.0078
R² = 0.999
Peak Area Ratio (11Del @ lesion site/9mer-5'P)
Molar Ratio (11Del @ lesion site/9mer-5'P)

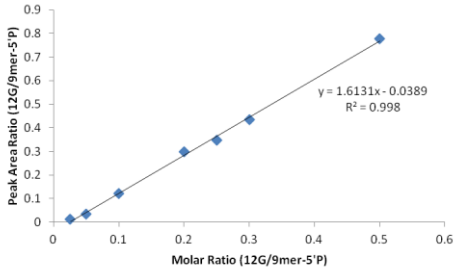
(P)


y = 0.9042x - 0.0118
R² = 0.993
Peak Area Ratio (12A/9mer-5'P)
Molar Ratio (12A/9mer-5'P)

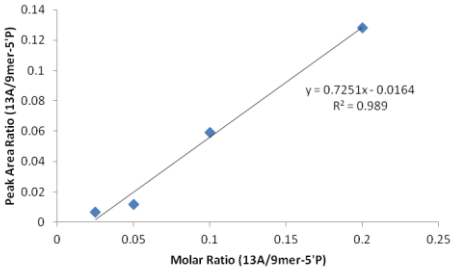
(Q)


y = 1.3275x - 0.0544
R² = 0.980
Peak Area Ratio (12A/9mer-5'P)
Molar Ratio (12A/9mer-5'P)

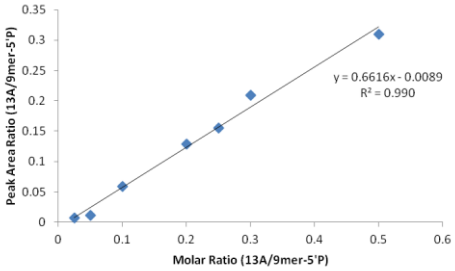
(R)


y = 1.6131x - 0.0389
R² = 0.998
Peak Area Ratio (12G/9mer-5'P)
Molar Ratio (12G/9mer-5'P)

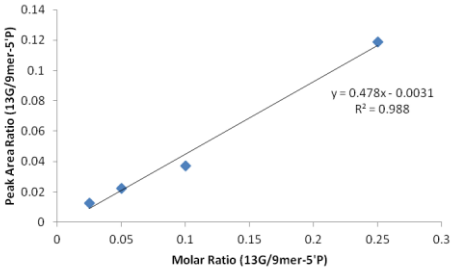
(S)


y = 0.7251x - 0.0164
R² = 0.989
Peak Area Ratio (13A/9mer-5'P)
Molar Ratio (13A/9mer-5'P)

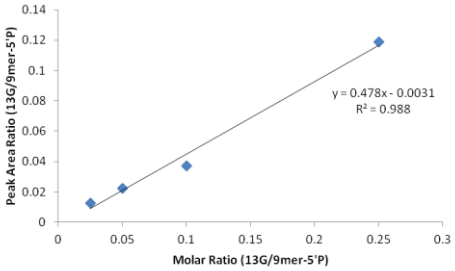
(T)


y = 0.6616x - 0.0089
R² = 0.990
Peak Area Ratio (13A/9mer-5'P)
Molar Ratio (13A/9mer-5'P)

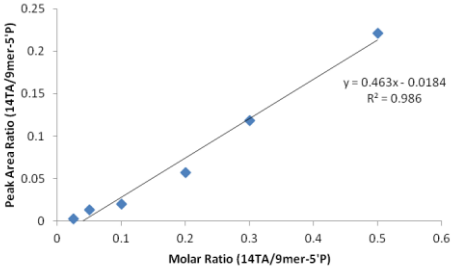
(U)


y = 0.478x - 0.0031
R² = 0.988
Peak Area Ratio (13G/9mer-5'P)
Molar Ratio (13G/9mer-5'P)

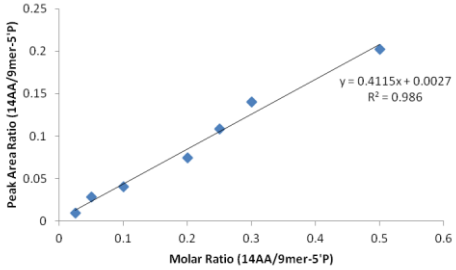
(V)


y = 0.478x - 0.0031
R² = 0.988
Peak Area Ratio (13G/9mer-5'P)
Molar Ratio (13G/9mer-5'P)

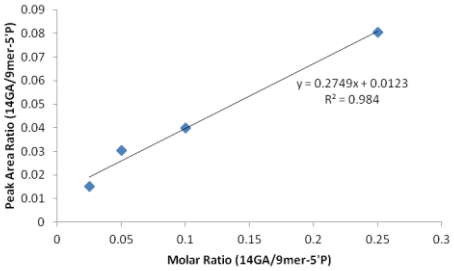
(W)


y = 0.463x - 0.0184
R² = 0.986
Peak Area Ratio (14TA/9mer-5'P)
Molar Ratio (14TA/9mer-5'P)

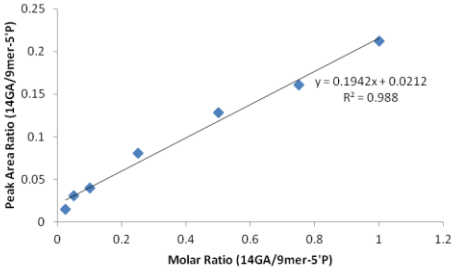
(X)


y = 0.4115x + 0.0027
R² = 0.986
Peak Area Ratio (14AA/9mer-5'P)
Molar Ratio (14AA/9mer-5'P)

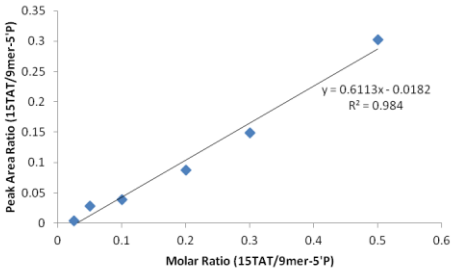
(Y)


Peak Area Ratio (14GA/9mer-5'P)
Molar Ratio (14GA/9mer-5'P)
y = 0.2749x + 0.0123
R² = 0.984

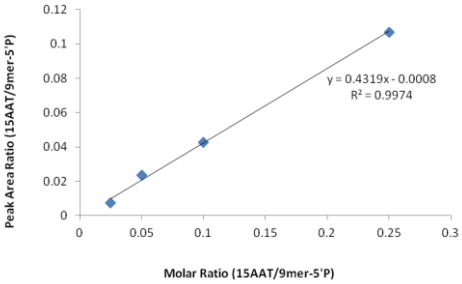
(Z)


Peak Area Ratio (14GA/9mer-5'P)
Molar Ratio (14GA/9mer-5'P)
y = 0.1942x + 0.0212
R² = 0.988

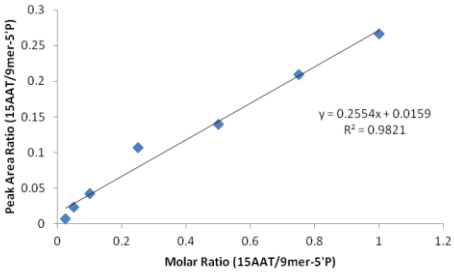
(AA)


Peak Area Ratio (15TAT/9mer-5'P)
Molar Ratio (15TAT/9mer-5'P)
y = 0.6113x - 0.0182
R² = 0.984

(BB)


Peak Area Ratio (15AAT/9mer-5'P)
Molar Ratio (15AAT/9mer-5'P)
y = 0.4319x - 0.0008
R² = 0.9974

(CC)


Peak Area Ratio (15AAT/9mer-5'P)
Molar Ratio (15AAT/9mer-5'P)
y = 0.2554x + 0.0159
R² = 0.9821

(DD)

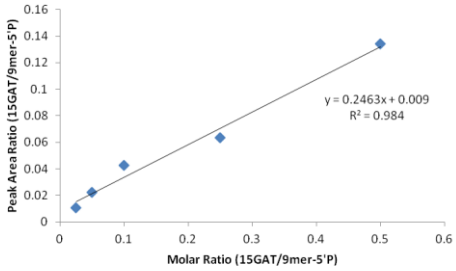
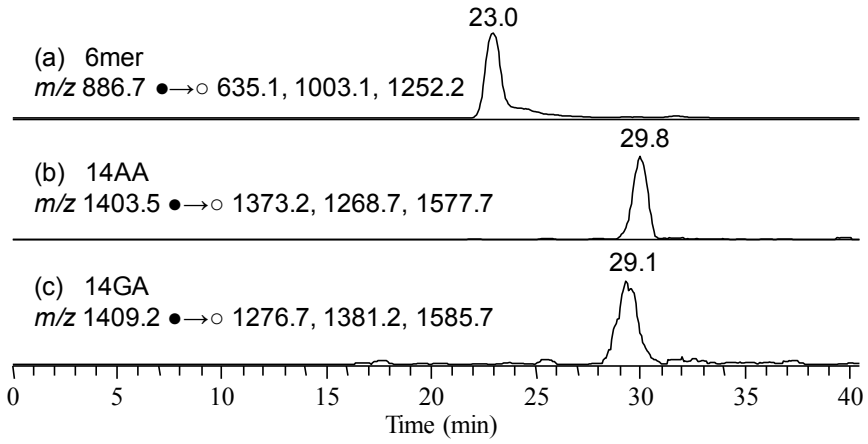

Peak Area Ratio (15GAT/9mer-5'P)
Molar Ratio (15GAT/9mer-5'P)
y = 0.2463x + 0.009
R² = 0.984

Figure B.7. Selected-ion chromatograms obtained from the LC-MS and MS/MS analysis of the hPol η-induced replication products for *N*3-EtdT-bearing substrate (A) and summary of percentages for each product (B). The replication products have been treated with restriction enzyme SacI and shrimp alkaline phosphatase, and 10-pmol of the replication mixture was injected for analyses.

(A)



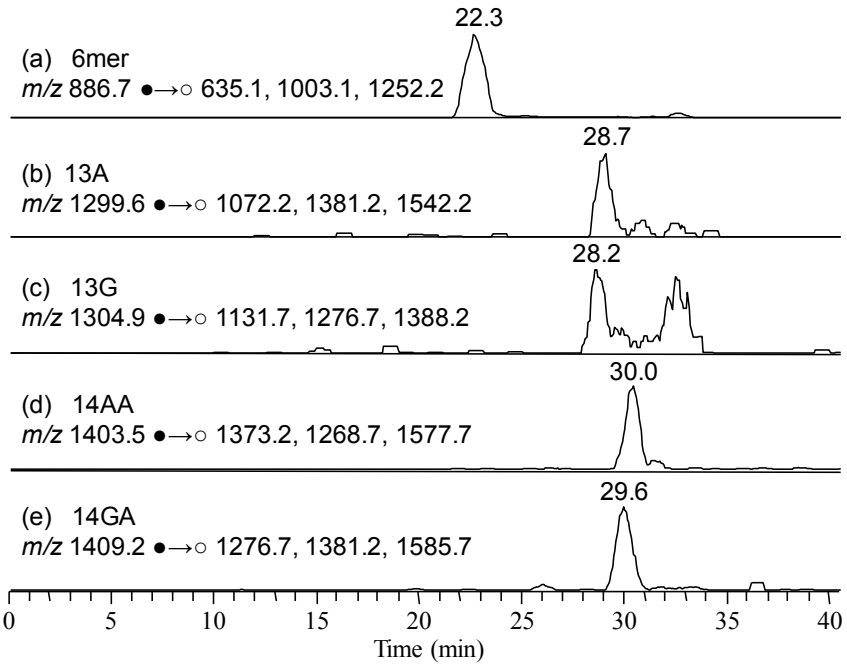
(B)

```
3′-CTAGCTCGAGTATCGXGCGGTA -5′
5′-          CATAGC        -3′   49%
5′-          CATAGCACGCCATA-3′   32%
5′-          CATAGCGCGCCATA-3′   19%
```

hPol η-induced Replication Products of N3-EtdT Substrate

Figure B.8. Selected-ion chromatograms obtained from the LC-MS and MS/MS analysis of the hPol η-induced replication products for O^4-EtdT-bearing substrate (A) and summary of percentages for each product (B). The replication products have been treated with restriction enzyme SacI and shrimp alkaline phosphatase, and 10-pmol of the replication mixture was injected for analyses.

(A)



(B)

```
3′-CTAGCTCGAGTATCGXGCGGTA -5′
5′-         CATAGC        -3′   26%
5′-         CATAGCACGCCAT -3′   2%
5′-         CATAGCGCGCCAT -3′   4%
5′-         CATAGCACGCCATA-3′   17%
5′-         CATAGCGCGCCATA-3′   51%
```

hPol η-induced Replication Products of O^4-EtdT Substrate

Table B.1. A summary of the percentages of products formed from hPol ι-mediated replication of dT-, O^2-EtdT-, $N3$-EtdT- and O^4-EtdT-containing substrates as determined by LC-MS/MS. The bases highlighted in bold represent those incorporated during primer extension. Listed under sequences are the SacI cleavage products bearing the extended portion of the primer.

Name	Sequence	dT	O^2-EtdT	$N3$-EtdT	O^4-EtdT
6mer unextended primer	5'-CATAGC-3'	22	50	16	35
7C +1	5'-CATAGC**C**-3'		1		6
7T +1	5'-CATAGC**T**-3'	17	38	17	15
7A +1	5'-CATAGC**A**-3'		4	2	28
7G +1	5'-CATAGC**G**-3'	51	5	64	16
8G +2	5'-CATAGC**GC**-3'	10	2		

Table B.2. A summary of the percentages of products formed from yPol ζ-mediated replication of dT-, O^2-EtdT-, *N*3-EtdT- and O^4-EtdT-containing substrates as determined by LC-MS/MS. The bases highlighted in bold represent those incorporated during primer extension. Listed under sequences are the SacI cleavage products bearing the extended portion of the primer.

Name	Sequence	dT	O^2-EtdT	*N*3-EtdT	O^4-EtdT
6mer unextended primer	5'-CATAGC-3'	29	96	93	26
7C +1	5'-CATAGC**C**-3'		2	1	
7T +1	5'-CATAGC**T**-3'			6	
7A +1	5'-CATAGC**A**-3'	1	2		
8A +2	5'-CATAGC**AC**-3'	2			
9G +3	5'-CATAGC**GCG**-3'				18
10A +4	5'-CATAGC**ACGC**-3'	20			
11A +5	5'-CATAGC**ACGCC**-3'	30			
12A +6	5'-CATAGC**ACGCCA**-3'	14			
12G +6	5'-CATAGC**GCGCCA**-3'				15
13A +7	5'-CATAGC**ACGCCAT**-3'	2			
13G +7	5'-CATAGC**GCGCCAT**-3'				41

Appendix C: Supporting Information for Chapter 4

Figure C.1. The product-ion spectrum of the $[M–4H]^{4-}$ ion (*m/z* 1317.4) of d(GCGCAAAXCTAGAGCTC) (X=S-N^2-CEdG). Negative-ion ESI-MS and a scheme outlining the fragments found in MS/MS are shown above the MS/MS.

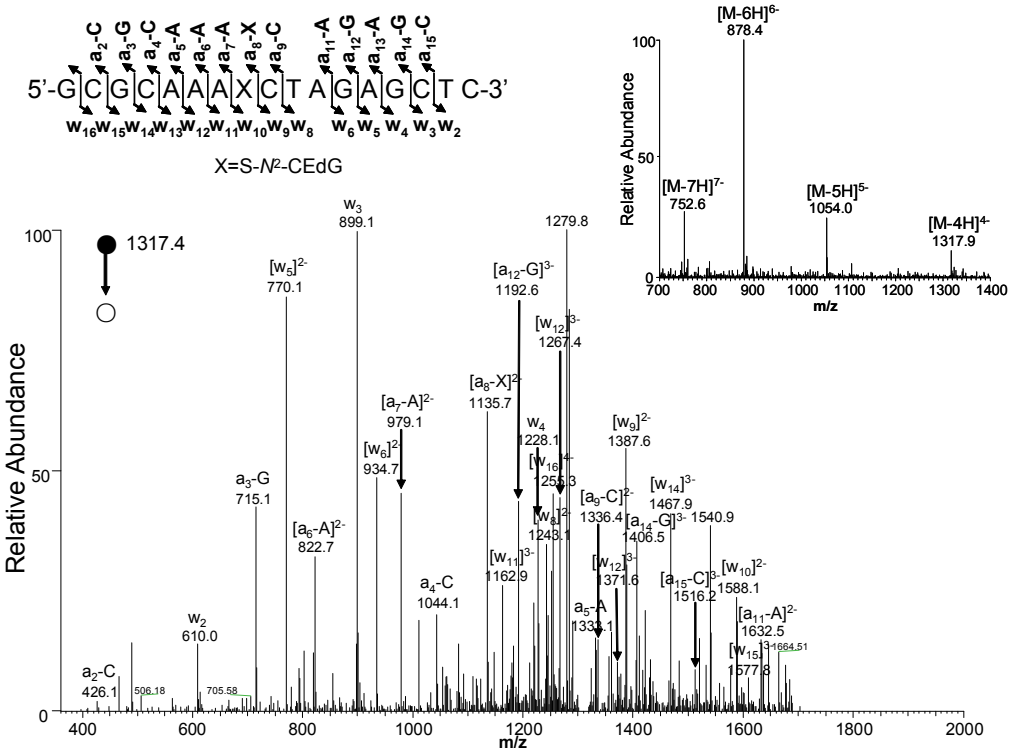


Figure C.2. The product-ion spectrum of the $[M–4H]^{4-}$ ion (*m/z* 1313.9) of d(GCGCAAAXCTAGAGCTC) (X=N^2-CMdG). Negative-ion ESI-MS and a scheme outlining the fragments found in MS/MS are shown above the MS/MS.

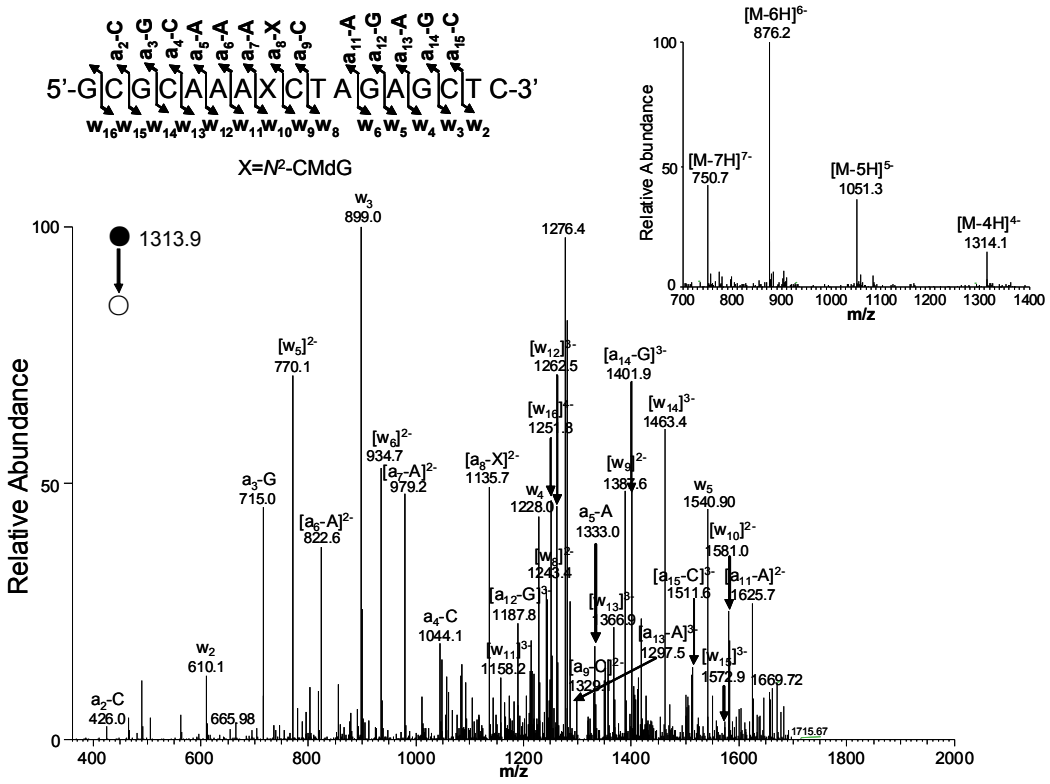


Figure C.3. LC-MS/MS for monitoring the restriction fragments of interest with a G→T (A) or G→A (B) mutation at the original N^2-CEdG and N^2-CMdG sites [i.e., d(GCAAATCTAGAGCT) (14mer-T) and d(GCAAAACTAGAGCT) (14-mer A)]. Shown in (a) and (b) the MS/MS of the $[M–4H]^{4-}$ ions (*m/z* 1066.4 and 1068.6) of these two ODNs.

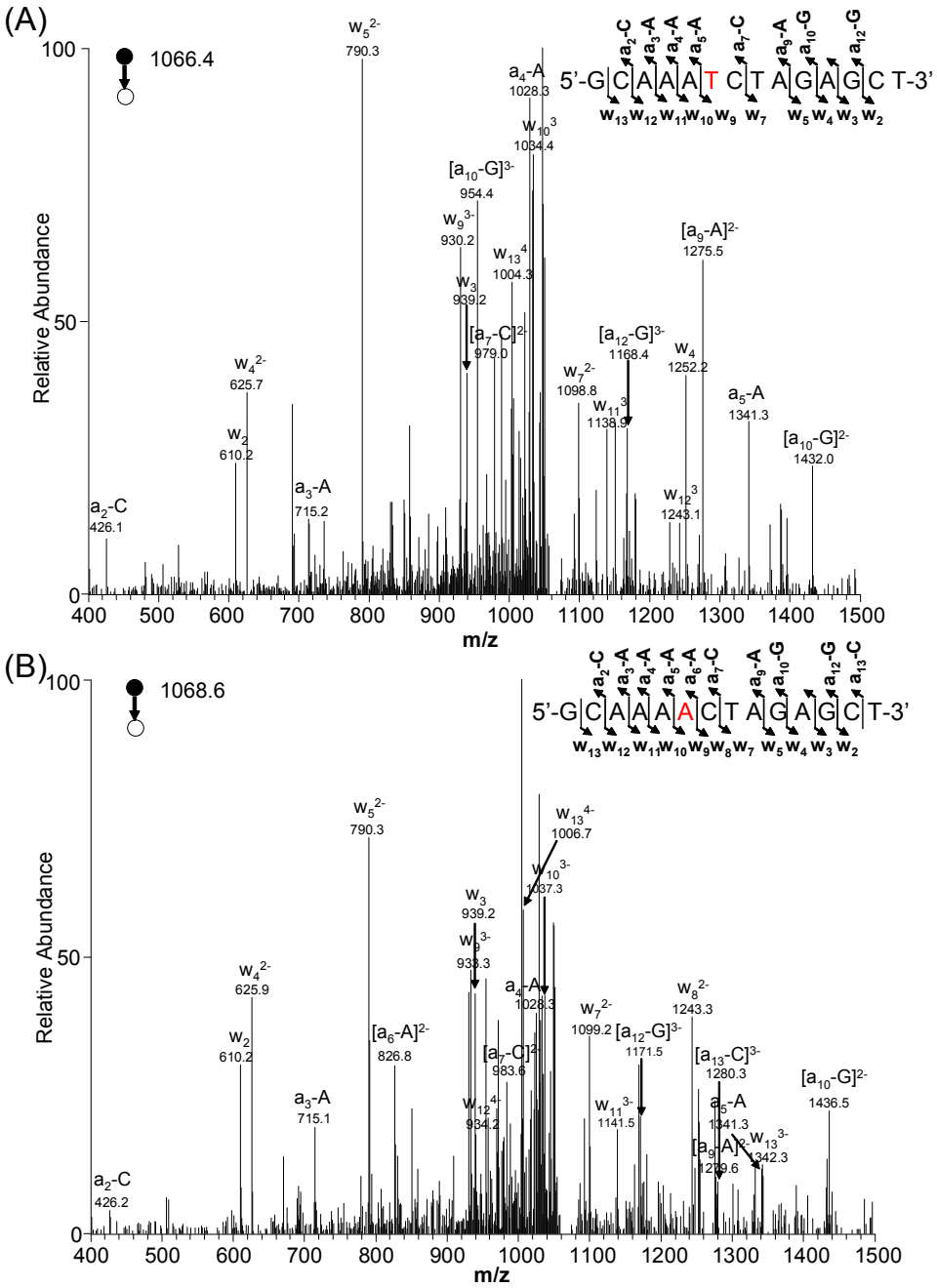
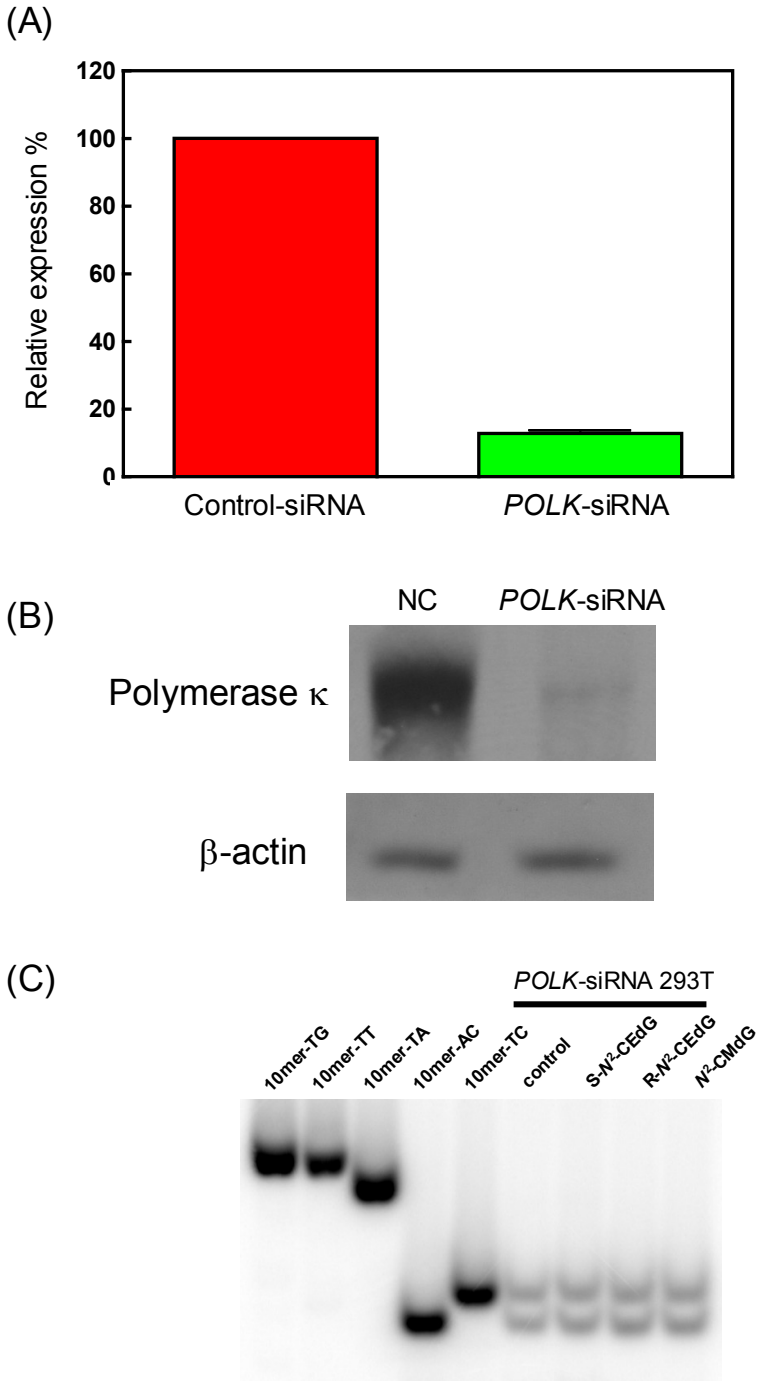


Figure C.4. Real-time qRT-PCR (A) and Western blot (B) analysis for monitoring the siRNA-induced knockdown of *POLK* gene in 293T cells and native PAGE (C) for assessing the mutation frequencies of N^2-CEdG and N^2-CMdG in the *POLK* siRNA-treated 293T cells.



Appendix D: Supporting Information for Chapter 5

Scheme D.1. Methods for the determination of cytotoxicity and mutagenicity of *c-Myb*-G4 sequences in *E. coli* cells.

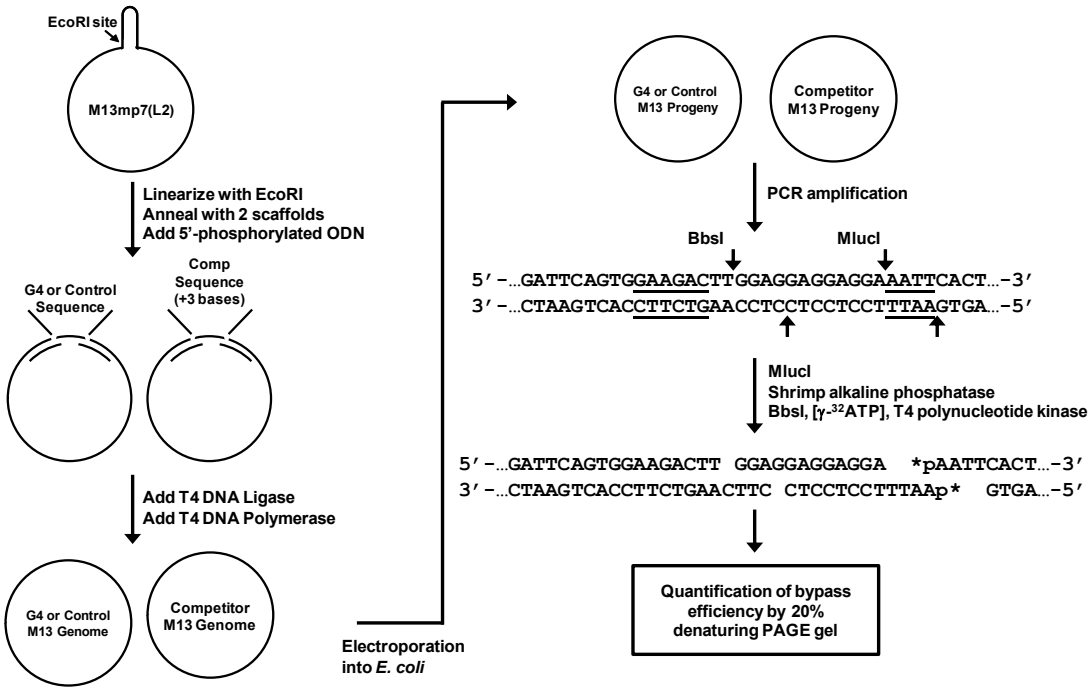


Table D.1. Sequences of oligodeoxynucleotides used in G-quadruplex experiments.

Name	Sequence
c-kit Sequences	
c-kit G4	5'-AGGGAGGGCGCTGGGAGGAGGG-3'
c-kit Control	5'-AGGGAGGGCTCTGTGAGGAGGG-3'
c-kit Competitor	5'-AGGGAGGGCTTGAGGAGGG-3'
c-kit Inserts	
c-kit G4	5'-GAAGACAAAGGGAGGGCGCTGGGAGGAGGG-3'
c-kit Control	5'-GAAGACAAAGGGAGGGCTCTGTGAGGAGGG-3'
c-kit Competitor	5'-GAAGACAAAGGGAGGGCTTGAGGAGGG-3'
c-kit Construction Scaffolds	
c-kit Scaffold 1	5'-TTGTCTTCCACTGAATCATGGTTCATAGC-3'
c-kit Scaffold 2	5'-AAAACGACGGCCAGTGAATTTCCTCC-3'
c-myb Sequences	
c-myb G4	5'-GGAGGAGGAGGA-3'
c-myb Control	5'-GCAGCAGCAGCA-3'
c-myb Competitor	5'-GCAGCACTGGCAGCA-3'
c-myb Inserts	
c-myb G4	5'-GAAGACTTGGAGGAGGAGGA-3'
c-myb Control	5'-GAAGACTTGCAGCAGCAGCA-3'
c-myb Competitor	5'-GAAGACTTGCAGCAGCAGCAGCA-3'
c-myb Construction Scaffolds	
c-myb Scaffold 1	5'-AAGTCTTCCACTGAATCATGGTTCATAGC-3'
c-myb Scaffold 2 (G4)	5'-AAAACGACGGCCAGTGAATTTCCTCC-3'
c-myb Scaffold 2 (Control/Competitor)	5'-AAAACGACGGCCAGTGAATTTGCTGC-3'